

10/803,704

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されて  
る事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed  
with this Office.

出 願 年 月 日            2 0 0 3 年 1 0 月    7 日  
Date of Application:

出 願 番 号            特 願 2 0 0 3 - 3 4 7 9 4 2  
Application Number:

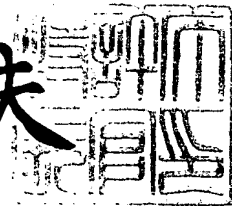
ST. 10/C] :            [ J P 2 0 0 3 - 3 4 7 9 4 2 ]

願            人            株式会社日立製作所  
Applicant(s):

2 0 0 4 年    3 月 2 5 日

特許庁長官  
Commissioner,  
Japan Patent Office

今 井 康 夫



BEST AVAILABLE COPY

CERTIFIED COPY OF  
PRIORITY DOCUMENT

出証番号    出証特 2 0 0 4 - 3 0 2 4 6 8 6

【書類名】 特許願  
【整理番号】 340301093  
【あて先】 特許庁長官殿  
【国際特許分類】 G06F 12/08  
G06F 03/06

【発明者】  
【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I  
D システム事業部内  
【氏名】 田村 美佐子

【発明者】  
【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I  
D システム事業部内  
【氏名】 竹内 久治

【発明者】  
【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I  
D システム事業部内  
【氏名】 古海 昇

【特許出願人】  
【識別番号】 000005108  
【氏名又は名称】 株式会社日立製作所

【代理人】  
【識別番号】 100095371  
【弁理士】  
【氏名又は名称】 上村 輝之

【選任した代理人】  
【識別番号】 100089277  
【弁理士】  
【氏名又は名称】 宮川 長夫

【選任した代理人】  
【識別番号】 100104891  
【弁理士】  
【氏名又は名称】 中村 猛

【手数料の表示】  
【予納台帳番号】 043557  
【納付金額】 21,000円

【提出物件の目録】  
【物件名】 特許請求の範囲 1  
【物件名】 明細書 1  
【物件名】 図面 1  
【物件名】 要約書 1  
【包括委任状番号】 0110323

**【書類名】 特許請求の範囲****【請求項 1】**

1 以上の情報要素を有する情報セットを発行する 1 又は複数の情報セットソースから前記情報セットの各々に含まれる情報要素を受信する受信手段と、

前記受信された情報要素の処理を行う情報処理手段と、

受信された 2 以上の情報セットに含まれる未処理又は処理中の複数の情報要素に基づいて、前記 2 以上の情報セット又は前記複数の情報要素の処理手順を決定するものであって、前記 2 以上の情報セットの処理時間長の平均に関する値が、前記複数の情報要素又は前記 2 以上の情報セットをその受信順序に従って処理するときの値以下になるような、前記受信順序とは異なる前記処理手順を決定する決定手段と

を備え、前記情報処理手段は、前記決定された処理手順に基づいて、前記未処理又は処理中の複数の情報要素の処理を開始する、  
情報処理装置。

**【請求項 2】**

前記決定手段は、前記複数の情報要素に基づいて、前記 2 以上の情報セットの各々について、処理に要すると期待される時間長に関する期待値を取得し、前記取得された期待値に基づいて、前記処理手順を決定する、

請求項 1 記載の情報処理装置。

**【請求項 3】**

情報要素に関する複数の情報要素属性の各々について、その情報要素属性を持つ情報要素の処理に要すると期待される時間長に関するサブ期待値を表すサブ期待値情報を記憶するサブ期待値記憶手段を更に備え、

前記決定手段が、前記複数の情報要素の各々の情報要素属性と、前記記憶されているサブ期待値情報とに基づいて、前記複数の情報要素の各々についての前記サブ期待値を取得し、前記取得されたサブ期待値を用いて、前記 2 以上の情報セットの各々についての前記期待値を取得する、

請求項 2 記載の情報処理装置。

**【請求項 4】**

前記決定手段は、前記取得されたサブ期待値が小さい情報要素又は情報セットほど早い順番で処理されるような前記処理手順を決定する、

請求項 3 記載の情報処理装置。

**【請求項 5】**

前記決定手段は、前記情報処理手段によって過去に処理された情報要素の属性とその情報要素の処理に要した時間長との履歴に関する情報に基づいて、前記 2 以上の情報セットの各々についての前記期待値を取得する、

請求項 2 記載の情報処理装置。

**【請求項 6】**

前記決定手段は、前記複数の情報要素についての複数通りの前記処理手順にそれぞれ対応した複数通りの前記平均に関する値のうち最小の値となるような前記処理手順を決定する、

請求項 1 記載の情報処理装置。

**【請求項 7】**

前記決定手段は、前記 2 以上の情報セットについて、前記処理時間長の最小と最大との差分を取得し、前記差分が所定値以上である場合に、前記処理手順を決定する処理を実行する、

請求項 1 記載の情報処理装置。

**【請求項 8】**

前記複数の情報要素に、前記情報処理手段によって処理中の情報要素が含まれているときは、前記決定手段は、その情報要素の処理に要する処理時間長のうちの残りの時間長に基づいて、前記処理手順を決定する、

請求項 1 記載の情報処理装置。

【請求項 9】

前記情報セットについての処理が完了する場合に、その処理が終了したことを意味する終了信号をその情報セットの発行元の情報セットソースに送信する終了信号送信手段を更に備え、

前記情報セットについての前記処理時間長は、前記情報セットソースが前記情報セットを発行してから前記終了信号を受信するまでのレスポンスタイムに関する時間長である、請求項 1 記載の情報処理装置。

【請求項 10】

前記決定手段は、前記受信された 2 以上の情報セットの受信状況が所定の状況である場合に、前記処理手順を決定する処理を実行する、請求項 1 記載の情報処理装置。

【請求項 11】

前記決定手段は、前記情報セット又は前記情報要素についての処理が開始されてから所定時点までの経過時間長を把握し、前記経過時間長が一定の時間長を超えたときに、前記処理手順を決定する処理を実行する、請求項 1 記載の情報処理装置。

【請求項 12】

物理的又は論理的な 1 又は複数の記憶デバイスに接続することができる記憶制御装置であって、

1 以上の情報要素を有する I/O 要求を発行する 1 又は複数の I/O 要求ソースから前記 I/O 要求の各々に含まれる情報要素を受信する受信手段と、

前記受信された I/O 要求に含まれている情報要素に基づいて、前記記憶デバイスからデータをリードし I/O 要求ソースに送信する、又は、前記 I/O 要求ソースからのデータを前記記憶デバイスにライトする処理を行う情報処理手段と、

受信された 2 以上の I/O 要求に含まれる未処理又は処理中の複数の情報要素に基づいて、前記 2 以上の I/O 要求又は前記複数の情報要素の処理手順を決定するものであって、前記 2 以上の I/O 要求のレスポンスタイムの平均に関する値が、前記複数の情報要素又は前記 2 以上の I/O 要求をその受信順序に従って処理するときの値以下になるような、前記受信順序とは異なる前記処理手順を決定する、請求項 1 記載の情報処理装置。

【請求項 13】

1 以上の情報要素を有する情報セットを発行する 1 又は複数の情報セットソースから前記情報セットの各々に含まれる情報要素を受信するステップと、

前記受信された情報要素の処理を行うステップと、

受信された 2 以上の情報セットに含まれる未処理又は処理中の複数の情報要素に基づいて、前記 2 以上の情報セット又は前記複数の情報要素の処理手順を決定するものであって、前記 2 以上の情報セットのレスポンスタイムの平均に関する値が、前記複数の情報要素又は前記 2 以上の情報セットをその受信順序に従って処理するときの値以下になるような、前記受信順序とは異なる前記処理手順を決定するステップとを有し、前記処理を行うステップでは、前記決定された要素処理手順に基づいて、前記未処理又は処理中の複数の情報要素の処理を開始する、情報処理方法。

【請求項 14】

1 以上の情報要素を有する情報セットを発行する 1 又は複数の情報セットソースから前記情報セットの各々に含まれる情報要素を受信するステップと、

前記受信された情報要素の処理を行うステップと、

受信された 2 以上の情報セットに含まれる未処理又は処理中の複数の情報要素に基づいて、前記 2 以上の情報セット又は前記複数の情報要素の処理手順を決定するものであって、前記 2 以上の情報セットのレスポンスタイムの平均に関する値が、前記複数の情報要素

又は前記 2 以上の情報セットをその受信順序に従って処理するときの値以下になるような、前記受信順序とは異なる前記処理手順を決定するステップと  
をコンピュータに実行させるためのコンピュータ読取可能なコンピュータプログラムであって、

前記処理を行うステップでは、前記決定された要素処理手順に基づいて、前記未処理又は処理中の複数の情報要素の処理を開始する、  
コンピュータプログラム。

## 【書類名】明細書

## 【発明の名称】情報処理装置及び方法

## 【技術分野】

## 【0 0 0 1】

本発明は、1以上の情報要素を有する情報セットを処理するための技術に関し、具体的には、例えば、E S C O N 或いは F I C O N 等のプロトコルに従って外部装置から受信する I / O 要求を処理するための技術に関する。

## 【背景技術】

## 【0 0 0 2】

1以上の情報要素を有する情報セットとして、例えば、コマンド又はデータである1以上のコマンド要素を有するコマンド要素セットがある。具体的な例としては、E S C O N (Enterprise Systems Connection)の通信プロトコルの下でやり取りされるコマンド要素セットであって、C C W チェインと呼ばれるものがある。

## 【0 0 0 3】

C C W チェインは、例えば、パーソナルコンピュータ等の上位装置と、データ記憶装置及びその記憶制御装置を持つ記憶システム（例えば R A I D (Redundant Array of Independent Inexpensive Disks) システム）とが E S C O N プロトコルに基づいて接続されている場合、上位装置から記憶システムに対して発行される I / O 要求である。

## 【0 0 0 4】

I / O 要求には、例えば、データ記憶装置からデータをリードするためのリードコマンドを含むもの（つまりリード要求）と、データ記憶装置に対しデータをライトするためのライトコマンドを含むもの（つまりライト要求）とがある。記憶システムにおける記憶制御装置は、キャッシュメモリを有しており、そのキャッシュメモリを介して、上位装置とデータ記憶装置との間でデータのリード及びライトが行われる。具体的に言えば、例えば、記憶制御装置は、上位装置からリード要求を受けた場合、リード要求において求められているリード要求データがキャッシュメモリに存在すれば（キャッシュヒットすれば）、そのリード要求データを上位装置に送信し、一方、それがキャッシュメモリに無ければ（キャッシュミスすれば）、データ記憶装置にアクセスしてリード要求データをリードしそれをキャッシュメモリに格納する処理（以下、これを「ステージング処理」と言う）を行ってから、そのリード要求データを上位装置に送信する。また、記憶制御装置は、ライト要求を受けた場合、そのライト要求に対応するキャッシュメモリ領域にターゲットデータが存在すれば（キャッシュヒットすれば）、そのライト要求に従うライト要求データをそのキャッシュメモリ領域に上書きし、その後、そのデータをデータ記憶装置に格納し、一方、上記ターゲットデータがキャッシュメモリ領域に無ければ（キャッシュミスすれば）、データ記憶装置にアクセスしてターゲットデータをリードしそれをキャッシュメモリ領域に格納するステージング処理を行ってから、そのライト要求データを上記キャッシュメモリ領域に上書きし、その後、その上書きしたデータをデータ記憶装置に送信する。

## 【0 0 0 5】

記憶制御装置は、I / O 要求について一連の処理が完了したら、E S C O N プロトコルに従い、処理が完了した旨を表す終了通知を上位装置に送信する。

## 【0 0 0 6】

このような記憶制御装置の性能を評価するための一つの評価項目として、上位装置が I / O 要求を発行してからその終了通知を受けるまでの時間（以下、「I / O レスポンスタイム」と言う）の長さというものがある。通常、I / O レスポンスタイムが短いほど、記憶制御装置の性能が良いとされる。一つの I / O 要求についての I / O レスポンスタイムは、一般に、キャッシュミスしたときはキャッシュヒットしたときに比べて長くなる。なぜなら、キャッシュメモリへのアクセスにかかる時間に比べて、磁気ディスクに代表されるデータ記憶装置へのアクセスにかかる時間の方が長いためである。

## 【0 0 0 7】

このレスポンスタイムの短縮を図るための技術として、例えば、特許文献1（特開2002-140233号公報）に開示されている技術がある。この技術によれば、例えば、記憶制御装置は、上位装置から受領した複数のI/O要求のうち、キャッシュヒットするI/O要求については、上位装置に対するデータ転送を実行し、キャッシュミスするI/O要求については、対象データをデータ記憶装置からキャッシュメモリに読み出す処理と、上位装置に対するデータ転送とを並行して実行する。この結果、複数のI/O要求の処理に対する記憶制御装置から上位装置への終了通知順序が、上位装置からのI/O要求受領順序とは異なる場合がある。

【0008】

【特許文献1】特開2002-140233号公報

【発明の開示】

【発明が解決しようとする課題】

【0009】

ところで、上述したI/Oレスポンスタイムの長さに基づいて記憶制御装置の能力を評価する際、発行された複数のI/O要求のI/Oレスポンスタイムの平均（以下、「平均レスポンスタイム」と言う）についての考察が必要であると考えられる。なぜなら、例えば、記憶制御装置が、上位装置から第1と第2のI/O要求を受信している場合、第1のI/O要求についてはI/Oレスポンスタイムが短くても、第2のI/O要求についてはI/Oレスポンスタイムが大分長くなってしまふのならば、全体として短いI/Oレスポンスタイムを提供したことにはならないからである。上述した特許文献1に開示の技術では、平均レスポンスタイムを考慮した処理は行われていない。

【0010】

上述した問題点は、上記のようなI/O要求に限らず、1以上の情報要素を持つ情報セットを受信し処理する他の装置にも存在し得る。また、その装置の能力を評価する際の評価項目として、平均レスポンスタイムに限らず、他の複数の処理に要する平均の処理時間長も有り得る。

【0011】

従って、本発明の目的は、複数の情報セットの処理に要する処理時間長の平均に関する値を小さくすることができるようにすることにある。

【課題を解決するための手段】

【0012】

本発明に従う情報処理装置は、1以上の情報要素を有する情報セットを発行する1又は複数の情報セットソースから前記情報セットの各々に含まれる情報要素を受信する受信手段と、前記受信された情報要素の処理を行う情報処理手段と、受信された2以上の情報セットに含まれる未処理又は処理中の複数の情報要素に基づいて、前記2以上の情報セット又は前記複数の情報要素の処理手順を決定するものであって、前記2以上の情報セットの処理時間長の平均に関する値が、前記複数の情報要素又は前記2以上の情報セットをその受信順序に従って処理するときの値以下になるような、前記受信順序とは異なる前記処理手順を決定する決定手段とを備え、前記情報処理手段は、前記決定された処理手順に基づいて、前記未処理又は処理中の複数の情報要素の処理を開始する。

【0013】

ここで、「情報要素」とは、情報セットに含まれる情報の要素であって、所定の単位、例えばフレーム単位で送受信される要素であり、具体例としては、コマンド又はデータである。

【0014】

また、「情報セット」とは、1以上の上記情報要素を含んだ情報要素グループであり、例えば、1以上のコマンド及び1以上のデータのうち少なくとも一方を含むコマンドデータセットである。また、情報セットに含まれる1以上の情報要素は、例えば、所定の順番でシーケンシャルに処理される必要があり、そのように処理されることで、1つの情報セットについての処理が完了するようになっている。そのため、例えば、決定手段が、2以

上の情報セット或いは上記複数の情報要素の処理手順を決定する際、本来は第1の情報要素よりも後に処理される第2の情報要素が第1の情報要素よりも先に処理されるような処理手順を決定することはない。

【0015】

また、「2以上の情報セットの処理時間長の平均に関する値」とは、その2以上の処理時間長の平均値、或いは、その平均値から取得可能な値（例えば合計値）である。

【0016】

また、前記2以上の情報セットの各々についての「処理時間長」は、例えば、所定時点（例えば現時点）からその情報セットについての処理の完了に関する時点までの時間長である。「処理の完了に関する時点」とは、例えば、処理完了を意味する終了通知を情報処理装置が情報セットソースに送信した時点、その終了通知を情報セットソースが受信しそれを認識した時点、その終了通知に応答して情報処理装置が所定の応答信号を送信した時点、又は、その応答信号を情報処理装置が受信し認識した時点のいずれかである。

【0017】

本発明に従う情報処理装置は、ハードウェア、ソフトウェア及びそれらの組合せのうちの少なくとも1つにより構築することが可能である。ソフトウェア及びそれとハードウェアの組合せでその情報処理装置を構築する場合、そのソフトウェアのためのコンピュータプログラムは、ディスク型ストレージや半導体メモリなどのプログラム記憶媒体を通じて、或いはインターネット等の通信ネットワークを通じてコンピュータにインストール又はロードすることができ、それにより、上述の情報処理装置を構築することができる。

【発明の効果】

【0018】

本発明に従う情報処理装置は、受信された2以上の情報セットに含まれる未処理又は処理中の複数の情報要素に基づいて、それら2以上の情報セットの処理時間長の平均に関する値が、複数の情報要素又は2以上の情報セットをその受信順序に従って処理するときの値以下になるような、受信順序とは異なる処理手順を決定し、それに基づいて、上記未処理又は処理中の複数の情報要素の処理が開始される。これにより、2以上の情報セットの処理時間長の平均に関する値を、単に処理対象の受信順序に従って処理されるときに値以下にすることができる。

【発明を実施するための最良の形態】

【0019】

以下、図面を参照して本発明の実施の形態について説明する。

【0020】

図1は、本発明の一実施形態に係る情報処理装置が適用された記憶システムの全体構成を示す。

【0021】

本実施形態における記憶システムには、1又は複数台の記憶装置60と、記憶制御装置10とが備えられる。

【0022】

記憶装置60は、データを記憶することができる装置であり、例えば、複数のハードディスクをそれぞれ備えた複数のディスクドライブ600、600、・・・をアレイ状に有するハードディスク群である（従って、この記憶システムは、いわゆるRAID（Redundant

Array of Independent Inexpensive Disks）システムである）。この記憶装置60の複数のディスクドライブ600、600、・・・における物理的な記憶領域上には、論理的な記憶領域である複数の論理ボリューム（Logical Unit）を設定することができる。

【0023】

記憶制御装置10は、接続ポート100、103と、チャネル制御部101、104と、I/O要求バッファ109、110と、コマンド制御部102、105と、サービスプロ

セッサ50と、共有メモリ106と、キャッシュメモリ108と、ディスク制御部107とを備えている。

#### 【0024】

接続ポート100、103は、パーソナルコンピュータ等の1又は複数の外部装置が記憶制御装置10に対し通信可能に接続するためのポートである。例えば、接続ポート100には、上位装置20が通信可能に接続され、別の接続ポート103には、ダイレクタ70を経由して上位装置30及び上位装置40と通信可能に接続される。これにより、記憶制御装置10は、各上位装置20、30、40から発行されたI/O要求（入出力要求）を接続ポート100又は103を介して受信することができる。

#### 【0025】

チャネル制御部101及び104は実質的に機能が同じなので101について代表的に説明する。チャネル制御部101は、ハードウェア回路、ソフトウェア、又はそれらの組み合わせで構成することができ、上位にある接続ポート100に接続されている上位装置20と通信可能に接続される。チャネル制御部101は、上位装置20から受けたI/O要求を、自分の下位に備えられているコマンド制御部102がアクセス可能なI/O要求バッファ109に格納する。また、チャネル制御部101は、上位装置20から受けたI/O要求に基づくデータを上位装置20に送信したり上位装置20から受信したりする。

#### 【0026】

I/O要求バッファ109及び110は実質的に機能が同じなので109について代表的に説明する。I/O要求バッファ109は、自分の上位に備えられているチャネル制御部101が上位装置20から受信したI/O要求を記憶することができるバッファであって、所定の内部メモリ（例えばRAM又はSDRAM等）上に確保されたものである。

#### 【0027】

コマンド制御部102及び105は実質的に機能が同じなので102について代表的に説明する。コマンド制御部102は、ハードウェア回路、ソフトウェア、又はそれらの組み合わせで構成することができ、上位に備えられているI/O要求バッファ109に格納されているI/O要求を読み出し、そのI/O要求に従う処理を行なう。

#### 【0028】

サービスプロセッサ50は、所定のデータソース（例えば外付式のハードディスク装置或いは可搬型の不揮発性メモリ）又はユーザから記憶制御装置10に種々の情報をするための入力処理装置であり、ハードウェア回路、ソフトウェア、又はそれらの組み合わせで構成することができる。具体的には、例えば、サービスプロセッサ50は、操作パネル等のユーザインタフェースであり、そのユーザインタフェースを介して、ユーザが、所定の情報、一例として、記憶制御装置10の保守に関する作業指示に関する情報、各種制御パラメータ、各種のチューニングパラメータ、及び後述する性能情報テーブル1060に関する情報を入力する。このサービスプロセッサ50から入力された各種情報は共有メモリ106に格納され、記憶制御装置10内の各制御部102、107等から必要に応じて参照又は更新される。

#### 【0029】

共有メモリ106は、例えば、不揮発性の半導体メモリ（例えばEEPROM）或いはハードディスクに仮想的に設けられたメモリであり、コマンド制御部102、105やディスク制御部107により共有されるものである。この共有メモリ106には、後述する性能情報テーブル1060と統計情報テーブル1061とが用意される（性能情報テーブル1060と統計情報テーブル1061の更新方法及び使用方法については後述する）。また、この共有メモリ106には、他の所定情報、例えば、I/O要求の処理に必要な制御テーブルが用意される。制御テーブルは、例えば、1以上の記憶装置60上における物理的又は論理的な記憶領域（以下、「データ記憶領域」と言う）毎に用意されるものであり、各制御テーブルには、キャッシュメモリ108に構築された1又は複数の仮想記憶領域（以下、「キャッシュスロット」と言う）のアドレスや、データ記憶領域のアドレスや、I/O要求の種別（例えば、リードとライトのどちらを要求するものか）等が記録され

る。キャッシュメモリ 108 上のキャッシュスロットや、各データ記憶領域に対するデータの読出しや書込みは、そのデータ記憶領域に対応した制御テーブルに記録されている情報に基づいて行なわれる。なお、I/O 要求には、所定の情報（例えば、データ記憶領域のアドレス）が含まれており、その情報から、その I/O 要求の処理のときに使用する制御テーブルを特定することができるようになっている。

#### 【0030】

キャッシュメモリ 108 は、例えば、揮発性又は不揮発性の半導体メモリ（例えば、SRAM 又は EEPROM）であり、コマンド制御部 102、105 やディスク制御部 107 により共有されるメモリである。キャッシュメモリ 108 には、記憶装置 60 からリードされたデータ、又は記憶装置 60 にライトするデータが一時的に格納される。具体的には、例えば、キャッシュメモリ 108 には、所定の制御部（例えばコマンド制御部 102、105）によって、一つのデータ記憶領域（例えば論理ボリューム）に対して一つのキャッシュスロットが設定され、そのキャッシュスロットには、そのキャッシュスロットに対応しているデータ記憶領域にライトされるデータ、又は、そのデータ記憶領域からリードされたデータが一時的に格納される。この記憶システムでは、リードされたデータ又はライトされるデータは、常に、このキャッシュメモリ 108 を介して、コマンド制御部 102、105 とディスク制御部 107 との間で受け渡される。

#### 【0031】

ディスク制御部 107 は、各記憶装置 60 毎に備えられており、その対応している記憶装置 60 を制御するものである。例えば、ディスク制御部 107 は、共有メモリ 106 上の制御テーブルに基づいて、1 以上の記憶装置 60 におけるデータ記憶領域（例えばターゲットの論理ボリューム）に対してデータの読出し又は書込みを行なう。また、ディスク制御部 107 は、例えば、論理アドレスを物理アドレスに変換することにより、論理ボリュームに対する I/O 要求を物理ディスクへの I/O 要求に変換することができても良い。

#### 【0032】

以上が、本実施形態における記憶システムについての概略的な説明である。なお、図示しないが、記憶制御装置 10 には、例えば、コマンド制御部 102、105 と、ディスク制御部 107 と、共有メモリ 106 と、キャッシュメモリ 108 とを相互に通信可能に接続するスイッチング制御部が備えられる。スイッチング制御部は、例えば、高速スイッチング動作によりデータ伝送を行う超高速クロスバススイッチ等のような高速バスとして構成することができる。コマンド制御部 102、105 と、ディスク制御部 107 と、共有メモリ 106 と、キャッシュメモリ 108 との間のデータやコマンドの授受は、図示しないスイッチング制御部を介して行われる。

#### 【0033】

以下、上述した記憶システムにおける I/O 要求の処理流れの概要を、上位装置 20 が I/O 要求を発行した場合を例に採り説明する。なお、その説明では、発行された I/O 要求がリード要求を示す場合とライト要求を示す場合とに分けて説明する。

#### 【0034】

(1) I/O 要求がリード要求を示す場合。

#### 【0035】

上位装置 20 から発行された I/O 要求を、接続ポート 100 を介してチャネル制御部 101 が受信し、チャネル制御部 101 が、その I/O 要求を I/O 要求バッファ 109 に格納する。コマンド制御部 102 は、その格納された I/O 要求を読出し、その I/O 要求とそれから特定される制御テーブルとに基づいて、その I/O 要求においてリード要求されているデータ（以下、「リード要求データ」と言う）がキャッシュメモリ 108 に存在するか否かの判別を行う。

#### 【0036】

その判別の結果が肯定的の場合、すなわち、リード要求データがキャッシュメモリ 108 に存在する場合（キャッシュヒットした場合）、コマンド制御部 102 は、キャッシュ

メモリ 108 からリード要求データを取得し、そのリード要求データを、チャンネル制御部 101 を介して上位装置 20 に送信する。

【0037】

一方、上記判別の結果が否定的の場合、すなわち、リード要求データがキャッシュメモリ 108 に存在しない場合（キャッシュミスした場合）、コマンド制御部 102 は、記憶装置 60 内のリード要求データをいったんキャッシュメモリ 108 にリードすることをドライブ制御部 107 に指示する。それにより、ディスク制御部 107 によって、記憶装置 60 からリード要求データがリードされてキャッシュメモリ 108 に格納されたならば、コマンド制御部 102 は、そのリード要求データをキャッシュメモリ 108 から取得して上位装置 20 に送信する。なお、この処理において、記憶装置 60 内のリード要求データをキャッシュメモリ 108 にリードする処理の間、上位装置 20 と記憶制御装置 10 との間の接続は切り離される。このように、キャッシュミスが起きた場合、リード要求データをキャッシュメモリ 108 にリードする処理にかかる時間と、切り離されている上位装置 20 に記憶制御装置 10 が再接続するのにかかる時間だけ、キャッシュヒットした場合にくらべて性能の劣化が生じる。

【0038】

(2) I/O 要求がライト要求を示す場合。

【0039】

上位装置 20 から、ライト対象のデータ（以下、「ライト要求データ」と言う）を含んだ I/O 要求を、接続ポート 100 を介してチャンネル制御部 101 が受信し、チャンネル制御部 101 が、その I/O 要求を I/O 要求バッファ 109 に格納する。コマンド制御部 102 は、その格納された I/O 要求を読み出し、その I/O 要求とそれから特定される制御テーブルとに基づいて、キャッシュメモリ 108 上の所定のキャッシュスロットにデータが存在するか否かの判別を行う。

【0040】

その判別の結果が肯定的の場合、すなわち、データが所定のキャッシュスロットに存在する場合（キャッシュヒットした場合）、コマンド制御部 102 は、上記読み出した I/O 要求に含まれているライト要求データを、上記所定のキャッシュスロット内のデータに上書きする。

【0041】

一方、上記判別の結果が否定的の場合、すなわち、データが所定のキャッシュスロットに存在しない場合（キャッシュミスした場合）、コマンド制御部 102 は、上記読み出した I/O 要求から特定されるデータ記憶領域からデータをいったんキャッシュメモリ 108 にリードすることをドライブ制御部 107 に指示する。それにより、ディスク制御部 107 によって、そのデータ記憶領域からデータがリードされて所定のキャッシュスロットに格納されたならば、コマンド制御部 102 は、上記所定のキャッシュスロットに格納されたデータに、上記読み出した I/O 要求に含まれているライト要求データを上書きする。

【0042】

このようにして、キャッシュメモリ 108 にライト要求データが書き込まれたときは、当該ライト要求は終了したものとして、記憶制御装置 10 から上位装置 20 に終了報告が返される。なお、キャッシュメモリ 108 にライト要求データが書き込まれた時点では、一般に、そのデータは記憶装置 60 内のドライブ 600、600、…には反映されておらず、万が一、記憶制御装置 10 の障害等によりキャッシュメモリ 108 上のデータが喪失した場合、データロスが発生してしまうが、このようなデータロスの回避のため、本実施形態の記憶制御装置 10 において、キャッシュメモリ 108 内のドライブ未反映データの 2 重化が行われても良いし、キャッシュメモリ 108 が不揮発性メモリであっても良い。

【0043】

以上が、本実施形態の記憶システムにおける I/O 要求の処理流れの概要である。

## 【0044】

以下、本実施形態の記憶システムについて、主要な部分について詳細に説明する。まず、その主要な部分の理解を助けるための説明を行う。

## 【0045】

上位装置20、30、40と記憶制御装置10との間は、所定の通信プロトコル、例えば、或るI/O要求の入出力通信フレームと別のI/O要求の入出力通信フレームとが混在して通信を行うことができるような通信プロトコルが実装されたI/Oインターフェースにより接続されている。その通信プロトコルは、例えば、データ又はコマンドの送信と応答といったインターロックを取ってから次のデータ又はコマンドが送信されるようなシーケンシャルな通信が行われる通信プロトコル（一例として、ESCON(Enterprise Systems Connection)プロトコル)であっても良いし、インターロックを取らずに上位装置がパイプライン的にデータ又はコマンドを送信することができる通信プロトコル（一例として、FICON(Fibre CONnection)プロトコル、或いはFC-SB2プロトコル(FIBRE CHANNEL Single-Byte Command Code Sets-2 Mapping Protocol))であっても良い。このような通信プロトコルの下で、上位装置20、30、40から発行されたI/O要求の処理が行われるときには、I/O要求に含まれるコマンド又はデータや、そのコマンド又はデータに対する応答は、フレームという形式で、上位装置20、30、40と記憶制御装置10との間でやり取りされる。

## 【0046】

I/O要求は、記憶装置60内のデータの読み取り(READ)及び記憶装置60内へのデータの書き込み(WRITE)等を要求するためのものであり、上述した種類の通信プロトコル（例えばFC-SB2プロトコル）の下では、I/O要求は、CCW（チャネル・コマンド・ワード）チェーンと呼ばれるコマンドセットである。

## 【0047】

CCWチェーンには、少なくとも1種類以上のコマンド要素が含まれており、具体的には、例えば、Define Extent（以下「DX」と表記）、Locate Record（以下「LOC」と表記）、リードコマンド、ライトコマンド等がある。DXコマンドは、レコードと呼ばれるI/O要求の対象データへのアクセス可否情報やアクセスモード等を指定するコマンドであり、I/O要求が発行されるときに初めに送信されるコマンドである。LOCコマンドは、I/O要求対象レコードの記憶装置60上の位置（例えばデータ記憶領域のアドレス）を示すシリンダ番号/トラック番号/レコード番号などの情報を示す。

## 【0048】

図2を参照して、このCCWチェーン（I/O要求）に対する処理流れの具体例を説明する。

## 【0049】

この例では、データのリードやライトはレコードと呼ばれる単位で行われ、レコードが複数個集まって、トラックという単位が構成される。そして、この例では、上位装置20から、或るトラックのレコード1、レコード2、レコード3をライトするライトコマンドが含まれたI/O要求70が発行されたものとする。そのため、I/O要求70は、コマンドが連結した形のコマンドチェーンとして発行され、そのコマンドチェーン70は、define extent/locate data/write(R1)/write(R2)/write(R3)である。このコマンドチェーン70において、define extent/locate dataまでのコマンドが、記憶装置60内のドライブ内におけるデータの位置付け命令であり、write(Rn:N=1,2,3)コマンドが、ライト要求データのライト命令である。

## 【0050】

まず、上位装置20から記憶制御装置10へ、コマンドチェーンに含まれる1以上のコマンド又はデータのうちの初めのコマンド(define extent)のフレームが発行される。その後、当該コマンドを受け付けた記憶制御装置10から、そのコマンドの応答として、コマンドレスポンスフレームが返却される。FC-SB2プロトコル等によれば、このコマンドレ

スポンスフレームを待つことなく、上位装置 20 は当該コマンドのパラメータデータをデータフレームとして送信することできるが、本例では、コマンドレスポンス、データの順序で記す。このパラメータデータを上位装置 20 から受けた記憶制御装置 10 は、上述した define extent コマンドに対する処理を実行する。

#### 【0051】

次に、上位装置 20 からは、次のコマンド「locate record」やデータのフレームが発行され、記憶制御装置 10 は、そのコマンドやデータについての処理を実行する。この二つのコマンド「define extent」及び「locate record」により、ライト要求データの格納位置情報が分かり、記憶制御装置 10 では、キャッシュヒットかキャッシュミスかの判定が行われる。以下、キャッシュヒットの場合を例に採り記述する。

#### 【0052】

上位装置 20 から、ライトレコードコマンドが送信される。続いてライト要求データであるレコード 1 のデータがデータフレームとして上位装置 20 から送信される。記憶制御装置 10 は、その更新データをキャッシュメモリ 108 に書き込む。同様に、記憶制御装置 10 は、レコード 2 及びレコード 3 についてのライトレコードコマンド及びライト要求データを受信し、それらについてもキャッシュメモリ 108 に書き込む。

#### 【0053】

レコード 3 についての更新データがキャッシュメモリ 108（或いは記憶装置 60 における所定のデータ記憶領域）に書き込まれた後、記憶制御装置 10 は、I/O 要求 70 についての処理が完了した旨を意味するステータスフレームを、上記コマンド発行元の上位装置 20 に送信する。ステータスフレームを受信した上位装置 20 は、ステータスフレームを受信したことの応答を意味するステータスアクセプトフレームを送信する。

#### 【0054】

以上が、I/O 処理 70 の処理についての流れである。なお、この I/O 要求 70 についての処理は、記憶制御装置 10 が当該ステータスアクセプトフレームを受信した時点で終了したこととなる。また、I/O 要求 70 についての I/O レスポンスタイム 72 は、上位装置 20 が I/O 要求 70 を発行した時点から上記ステータスフレームを受信したことを認識した時点までの時間長である。また、図 2 は、I/O 要求がデータのライトを要求するものである場合のデータ通信のフローであるが、I/O 要求がデータのリードを要求するものである場合は、リード要求データの送信方向がライト要求データの送信方向と逆となる。また、図 2 は、キャッシュヒットした場合のフローであるが、キャッシュミスした場合は、記憶制御装置 10 は上位装置 20 に一旦接続の切離しを要求し、記憶装置 60 からキャッシュメモリ 108 へのデータの読み込みが完了した時点で、記憶制御装置 10 が上位装置 20 に再接続を要求することを除いては概ね同様のフローとなる。

#### 【0055】

さて、次に、複数の I/O 要求がある場合のレスポンスタイムについて考察する。

#### 【0056】

記憶制御装置 10 が上位装置 20 からの I/O 要求を受け付け、要求に応じた処理を行なう際に複数の I/O を多重で処理する場合、各々の処理を行なうスケジュールにより接続ポートの稼働率、プロセッサの稼働率を上げることが可能であり、その結果、単位時間あたりに処理可能な I/O 数は複数の I/O 要求を混在させずに一つ一つ処理した場合に比べて向上すると考えられる。しかし、一方で、多重処理された複数の I/O 要求それぞれに対するレスポンスタイムに着目すると、それぞれの I/O を多重処理せずに 1 つの I/O 要求ずつ処理を行なった方が、レスポンスタイムは短くなる場合がある。このケースの一例を図示したのが図 3 である。

#### 【0057】

この図 3 では、2 つの I/O 要求 [1]、[2] を、[1][2] の順に 1 つの I/O 要求ずつ処理をシーケンシャルに行った場合と、2 つの I/O 要求のフレームを混在させて行った場合をモデル化している。この図 3 によれば、シーケンシャルに実行しているケースに比べ、

2つのI/O要求のフレームを混在しているケースでは、プロセッサ（例えばコマンド制御部102又は105を含むプロセッサ）の空き時間が少なく、そのプロセッサや接続ポート100、103の稼働率は上がる。しかし、I/O要求[1]のレスポンスタイムについて着目すると、[1][2]の順序で1つのI/O要求ずつ処理をシーケンシャルに行った方が短くなっている。

#### 【0058】

このように、或る一つのI/O要求[1]の処理中に違うI/O要求[2]のフレームをすきまなく混在させ、また、双方向にフレームの送受信を行なうことによって、処理可能なI/O要求の数を増進できることが期待できる。しかし、複数のI/O要求の処理を行なうとはいえ、各I/O要求に含まれるコマンド単位或いはデータ単位で見れば、単一のI/O要求の処理の場合と実質的に変わらない。ただし、複数のI/O要求に対するフレームの各々が混在して送受信可能であるため、接続ポート100、103の空き時間を縮小させることは可能である。これにより、単位時間当りに処理可能なI/O要求の総数は、単一のI/O要求処理しか行えないプロトコルに比べて向上することが期待できる。しかし、同じコマンドからなるひとつのI/O要求に着目して、単一I/O処理プロトコルと多重I/O処理プロトコル両方で処理した場合のレスポンスタイムについて考えると、前述した図3に示すような結果を得ることができる。すなわち、同じコマンドからなる一つのI/O要求においては、単一I/O処理プロトコルによって処理された場合のレスポンスタイムの方が、多重I/O処理プロトコルによって処理された場合のレスポンスタイムより短くなる。勿論、どんなケースでも図3に例示した結果が同じとなるわけではないが、一般に多重の度合いが高い設定になるほど、ひとつのI/O要求に着目したI/Oレスポンスタイムの劣化が見られる。しかし、多重I/O処理では、ポートおよび処理プロセッサの稼働率が単一I/O処理に比べて向上することから、単位時間に処理可能なI/O要求の数は多くなり、システムにおけるスループット性能はこちらの方が優れている。つまり、情報処理システムにおける性能を測るものさしとして、I/Oレスポンスタイムの短縮とスループットの向上という2つが挙げられるであろうが、どちらを重視するかはシステム構成、運用形態などによってまちまちであり、特にバッチ処理かオンライン処理かが重要である。例えば、前者の場合はスループット重視、後者の場合はI/Oレスポンスタイムが重視される。本実施形態では、このI/Oレスポンスタイムについて、複数のI/O要求の平均レスポンスタイムに着目している。

#### 【0059】

なお、I/Oレスポンスタイムとは、前述したように、上位装置がI/O要求を発行してからその終了通知を受けるまでの時間長であるが、詳細には、図4からわかるように、（1）上位装置がI/O要求を発行してから記憶制御装置に到着するまでの時間長、（2）記憶制御装置においてI/O要求の処理が待ち状態になっている時間長、（3）記憶制御装置においてI/O要求が処理されている時間長、及び（4）記憶制御装置が終了通知を発行してから上位装置が受信するまでの時間長の合計である。

#### 【0060】

さて、次に、複数のI/O要求の平均レスポンスタイムについて考察し、その内容の一例を図4に示す。なお、この例では、I/O要求[3]、[4]、[5]が、上位装置20から略同時に発行されたと仮定する。そして、3つのI/O要求は、それぞれ3つのコマンドa、b、cからなるものとする。すなわち、I/O要求[3]は、コマンド[3]-a、[3]-b、[3]-cを含んでおり、I/O要求[4]、[5]についても実質的に同様であるものとする。また、図4に例示したケースでは、全てのコマンドa、b、cについて、処理にかかる時間が同一であるものとする。各コマンドの実行にかかる時間を $T_c$ とする。そのため、全てのI/O要求は3つのコマンドa、b、cから成るので、各I/O要求毎の処理に要する処理時間長は3 $T_c$ である。また、説明を分かりやすくするために、あるI/O要求が有するコマンドの処理が終了してから別のI/O要求が持つコマンドの処理を開始する際のオーバーヘッド（つまり、処理対象のI/O要求を切り替える際のオーバーヘッド）は省略することとする。

## 【0061】

この図4に示すように、ケース『3-A』では、I/O要求[3]、I/O要求[4]、I/O要求[5]の順序で処理が行われる。この場合、3つのI/O要求[3]、[4]、[5]のI/Oレスポンスタイムは、それぞれ、 $3T_c$ 、 $6T_c$ 、 $9T_c$ となる。従って、3つのI/O要求[3]、[4]、[5]についての平均レスポンスタイムは $6T_c$ となる。

## 【0062】

一方、同図において、ケース『3-B』では、I/O要求[3]のコマンド[3]-a、I/O要求[4]のコマンド[4]-a、I/O要求[5]のコマンド[5]-a、I/O要求[3]のコマンド[3]-b、I/O要求[4]のコマンド[4]-b、I/O要求[5]のコマンド[5]-b、I/O要求[3]のコマンド[3]-c、I/O要求[4]のコマンド[4]-c、I/O要求[5]のコマンド[5]-cの順で処理が行われる。着目すべき点は、各I/O要求についてのI/Oレスポンスタイムである。3つのI/O要求の処理が終了する時刻はケース『3-A』と同じであるにもかかわらず、それぞれのI/Oレスポンスタイムは、I/O要求[3]については $7T_c$ 、I/O要求[4]については $8T_c$ 、I/O要求[5]については $9T_c$ となり、それ故、それら3つのI/O要求の平均レスポンスタイムは、 $8T_c$ となる。

## 【0063】

なお、上記の説明では、平均レスポンスタイムについて考察した内容を理解し易くするため、処理対象のI/O要求の切り替え等に伴うオーバーヘッドは省略したが、このオーバーヘッドを加味すると、ケース『3-A』とケース『3-B』の平均レスポンスタイムの差はさらに顕著になると考えられる。図4の例は、複数のI/O要求のコマンド処理を混在させることにより、平均レスポンスタイムが劣化してしまう場合のモデルである。

## 【0064】

次に、複数のI/O要求の処理を行う場合、その実行順序による平均レスポンスタイムの違いについて、図5を使用して説明する。図5では、I/O要求[6]、[7]を例に説明する。なお、この例では、I/O要求[6]は、I/O要求[7]に比べて2倍の処理時間長がかかるものと仮定し、それ故、I/O要求[6]、[7]の処理にかかる処理時間長をそれぞれ $2T$ 、 $T$ であるものとする。また、この例では、図4を参照して説明した例と同様、I/O要求は上位装置20より略同時に発行されたものとし、且つ、理解の容易化のため、処理対象のI/O要求の切り替え等に伴うオーバーヘッドは省略する。

## 【0065】

ケース『4-A』では、I/O要求[6]、I/O要求[7]の順序で処理が実行される。このため、2つのI/O要求の平均レスポンスタイムは、 $2.5T$ である。

## 【0066】

一方、ケース『4-B』では、I/O要求[7]、I/O要求[6]の順序で処理が実行される。このため、2つのI/O要求の平均レスポンスタイムは $2T$ となり、ケース『4-A』における順序でI/O要求の処理が行われた場合よりも、平均レスポンスタイムは向上する。

## 【0067】

以上、I/O要求のレスポンスタイムの幾つかの例について説明した。1つ目の例、すなわち、図3を参照して説明した例では、複数のI/O要求のコマンド/フレーム処理を混在させることで、プロセッサ（例えばコマンド制御部102又は105を有するプロセッサ）の稼動効率が上がり、スループットの向上が望める反面、1つのI/O要求に着目するとレスポンスタイムが劣化する場合があることを示した。2つ目の例、すなわち、図4を参照して説明した例では、複数のI/O要求の平均レスポンスタイムに着目し、I/O要求を一つずつシーケンシャルに実行した場合と、複数I/O要求のコマンド/フレーム処理を混在させた場合とで、平均レスポンスタイムが異なる場合があることを示した。3つ目の例、すなわち、図5を参照して説明した例では、2つ目の例と同様、複数のI/O要求の平均レスポンスタイムに着目しているが、複数のI/O要求のコマンド/フレームの混在ではなく、複数のI/O要求の実行順序によって平均レスポンスタイムが変動する例を示した。

**【 0 0 6 8 】**

以上の例から分かるように、2 以上の I / O 要求の平均レスポンスタイムは、2 以上の I / O 要求に含まれる複数のコマンド又はデータをどのような順番で処理するかによって異なる場合もあれば、2 以上の I / O 要求そのものをどのような順番で処理するかによって異なる場合もある。

**【 0 0 6 9 】**

以上が、本実施形態の主要な部分の理解を助けるための説明である。以下、本実施形態の主要な部分について説明する。なお、その説明では、上位装置 2 0、3 0、4 0 との間の通信プロトコルは、或る I / O 要求の入出力通信フレームと別の I / O 要求の入出力通信フレームとが混在して通信を行うことができる通信プロトコル、例えば、FC-SB2 プロトコルで接続されているものとする。それ故、記憶制御装置 1 0 は、ESCON プロトコルと異なり、一つの I / O 要求の処理の実行中に、上位装置と記憶制御装置 1 0 と間の論理的な接続パス（以下、「論理パス」と表記）を占有することをせず、同一論理パス上で同時に複数のデータ記憶領域（例えば論理ボリューム）に対する I / O 要求の処理を実行することができる。また、上位装置 2 0 が、I / O 要求中のライト又はリードコマンドを記憶制御装置 1 0 に送信後、記憶制御装置 1 0 からコマンド応答を待つことなく、待つ必要もなく、ライト要求データを送信することが可能であり、さらに、上位装置は、当該コマンドに対するステータスフレームを受信しなくても、次のコマンド処理に移り、別の I / O 要求の発行が可能である。

**【 0 0 7 0 】**

図 6 は、記憶制御装置 1 0 の共有メモリ 1 0 6 に用意される性能情報テーブル 1 0 6 0 を示す。

**【 0 0 7 1 】**

性能情報テーブル 1 0 6 0 には、I / O 要求に含まれているコマンド又はデータに関する複数のコマンド／データ属性の各々について、その属性を持つコマンド又はデータの処理に要すると期待される時間長に関する期待値が記録されている（期待値は、時間長それ自体を意味するものであっても良いし、その時間長に関する別の意味を表すものであっても良い）。この性能情報テーブル 1 0 6 0 に記録される情報は、人間がサービスプロセス 5 0 を使用して入力した情報であっても良いし、記憶制御情報 1 0 が後述する統計情報に基づいて自動的に登録するものであっても良い。

**【 0 0 7 2 】**

性能情報テーブル 1 0 6 0 には、コマンド処理時間サブテーブル 1 0 6 0 A と、データ転送処理時間サブテーブル 1 0 6 0 B とが含まれている。

**【 0 0 7 3 】**

コマンド処理時間サブテーブル 1 0 6 0 A には、I / O 要求に含まれているコマンドに関する複数のコマンド属性（例えばコマンド種別）の各々について、その属性を持つコマンドの処理に要する期待値が記録されている。この期待値は、例えば、コマンドの処理を開始した時点からその処理が完了し所定の信号を上位装置に送信する時点までの時間長に関するものである。

**【 0 0 7 4 】**

データ転送処理時間サブテーブル 1 0 6 0 B には、I / O 要求に含まれているデータに関する複数のデータ属性（例えば転送種別及び転送長）の各々について、その属性を持つコマンドの処理に要する期待値が記録されている。転送種別が「リード処理」であるデータについての期待値は、例えば、キャッシュメモリ 1 0 8 からデータをリードして上位装置に送信するまでの時間長に関するものである。また、転送種別が「ライト処理」であるデータについての期待値は、例えば、I / O 要求バッファ 1 1 0 からデータを読出してキャッシュメモリ 1 0 8 にし、そのデータがキャッシュメモリ 1 0 8 から記憶装置 6 0 に格納されるまでの時間長に関するものである。

**【 0 0 7 5 】**

この図 6 に示した性能情報テーブル 1 0 6 0 によれば、例えば、或る I / O 要求が、リ

ード要求であって、且つ、コマンド種別[1]のコマンドと、コマンド種別[2]のコマンドと、データ長 a のデータとから構成されている場合は、その I/O 要求についての期待値の合計（換言すれば、その I/O 要求の処理にかかると予測される時間長）は、 $T_1 + T_2 + T_a$  である。

#### 【0076】

以上が、性能情報テーブル 1060 についての説明である。なお、言うまでも無いが、性能情報テーブル 1060 に記録される情報の種類は、図 6 に示したものに限られない。例えば、情報性能テーブル 1060 には、キャッシュヒットしたときとキャッシュミスしたときとに分けて、種々の期待値が記録されていても良い。また、コマンド処理時間サブテーブル 1060A に記録されている期待値、或いは、データ転送処理時間サブテーブル 1060B において転送種別が「リード処理」の期待値には、記憶制御装置 10 から送出されたデータが上位装置 20、30 又は 40 に到達するまでの時間長も含まれていても良い。その時間長は、上位装置 20、30 及び 40 の全てについて共通の値であっても良いし、上位装置 20、30 及び 40 毎に固有の値であっても良い。

#### 【0077】

次に、格別図示しないが、統計情報テーブル 1061 について説明する。

#### 【0078】

統計情報テーブル 1061 には、記憶制御装置 10 において過去に行われた I/O 要求処理に関する履歴に基づく統計情報が格納される。統計情報としては、例えば、コマンド又はデータに関する複数の属性の各々について、処理にかかった時間長の統計（例えば平均）を含んだ情報がある。

#### 【0079】

この統計情報テーブル 1061 には、例えば、以下のような流れで統計情報が格納される。

#### 【0080】

例えば、記憶制御装置 10 における各構成要素（例えばコマンド制御部 102 及び 105）は、I/O 要求中のコマンド又はデータの処理が行われたときは、その都度又は所定のタイミングで（例えば定期的に）、そのコマンド又はデータに関する履歴を所定の履歴格納領域（例えば不揮発性のメモリ）に登録する。登録される履歴は、例えば、過去一定期間における履歴である。

#### 【0081】

記憶制御装置 10 における所定の構成要素（例えばチャネル制御部及びコマンド制御部を有する各マイクロプロセッサ）は、所定のタイミングで（例えば定期的に）、上記履歴格納領域に書かれている履歴情報に基づいて、例えば過去一定期間における I/O 要求処理をトレースし、そのトレースの結果に基づいて、I/O 要求中のコマンド又はデータの属性毎に、そのコマンド又はデータの処理にかかった時間長の統計値を統計情報テーブル 1061 に格納する。

#### 【0082】

なお、記憶制御装置 10 における所定の構成要素（例えばチャネル制御部及びコマンド制御部を有する各マイクロプロセッサ）は、この統計情報テーブル 1061 に格納された上記時間長に基づいて、適宜に（例えば定期的に）、性能情報テーブル 60 の内容を更新しても良い。具体的には、例えば、上記所定の構成要素は、この統計情報テーブル 1061 から各エントリ項目（コマンド又はデータ属性）に対応した時間長を読み出し、性能情報テーブル 60 上のそのエントリ項目に対応した期待値に、上記読み出した時間長を上書きすることで、性能情報テーブル 1060 の内容を更新しても良い。

#### 【0083】

次に、図 7 を参照して I/O 要求バッファ 109 及び 110 について説明する。109 及び 110 は、構成的には実質的に同一なので、I/O 要求バッファ 109 について代表的に説明する。

#### 【0084】

I/O要求バッファ109は、I/O要求処理を行うプロセッサ（具体的には、チャンネル制御部101及びコマンド制御部102を含むプロセッサ）のバッファであり、同一のプロセッサ内にあるチャンネル制御部102およびコマンド制御部102よりアクセス可能である。このI/O要求バッファ109には、N個のI/O要求の各々について、バッファ領域が用意されており、各I/O要求のコマンド又はデータフレームは、そのI/O要求に対応したバッファ領域に格納することができるようになっている。また、I/O要求バッファ109には、現在このバッファ109に登録されているI/O要求の数を表すI/O要求数情報を登録することができるようになっている。また、このI/O要求バッファ109には、各I/O要求毎に、実行順序情報（例えば、処理される順位を表す情報）を登録することができるようになっている。

#### 【0085】

なお、I/O要求バッファ109の形態は、この図7に例示したように、I/O要求毎にバッファ領域が用意される形態に限られない。例えば、I/O要求バッファ109は、チャンネル制御部101が受け付けたコマンド又はデータフレームが、どのI/O要求に含まれるものであるかに関わらず、先着順に格納される形態（例えばFIFOの形態）でも良い。

#### 【0086】

上位装置20は、特定の通信プロトコルの下、発行する各I/O要求について、初めから或いは所定のコマンド（例えばlocate recordコマンド）に対する応答を受けた以後、そのI/O要求に含まれるコマンド又はデータを任意のタイミングで任意の数だけ記憶制御装置10に送信することができる。そのため、I/O要求バッファ109には、各I/O要求について、未処理状態のデータ又はコマンドが1つしか存在しない場合もあれば、2つ以上存在する場合もある。

#### 【0087】

チャンネル制御部101は、当該バッファ109の空きサイズや領域等を管理しており、上位装置20から最新のI/O要求を受けたときは、未だ何らコマンド又はデータが格納されていないバッファ領域にそのI/O要求中のコマンド又はデータを格納し、一方、受信済みのI/O要求における新たなコマンド又はデータフレームを受信したときは、そのI/O要求に対応したバッファ領域にそのコマンド又はデータフレームを追加する。また、チャンネル制御部101は、I/O要求バッファ109に登録されているI/O要求の数を把握し、それを表すI/O要求数情報をそのバッファ109に登録する。また、チャンネル制御部101は、その登録されているI/O要求の数が変化する都度に、バッファ109に登録されているI/O要求数情報を更新する。また、チャンネル制御部101は、バッファ109に残すべきデータや消去すべきデータを管理しても良い。具体的には、例えば、チャンネル制御部101は、最後まで処理が完了していない（例えば、処理終了を意味するステータスフレームが送信されていない或いはそれに対する応答を受信していない）I/O要求については、たとえ処理済みのコマンドやデータがあっても消去せず、全ての処理が完了した段階で、そのI/O要求に含まれているコマンド及びデータを全て消去するようにしても良い。このように、I/O要求に含まれるデータやコマンドはそれについての処理が完了するまでバッファ109に残るので、後述するように、I/O要求単位で後述の要求処理順序を決定することができる。

#### 【0088】

コマンド制御部102は、後に詳述する方法で、上位装置20から受信した2以上のI/O要求をどのような順序で処理するか（或いは、2以上のI/O要求に含まれる複数のデータ又はコマンドをどのような順序で処理するか）を決定する。そして、コマンド制御部102は、決定した順序（以下、「要求処理順序」と言う）に基づき、各I/O要求の実行順序情報（例えば処理順位）をI/O要求バッファ109（つまり、コマンド制御部102を持つマイクロプロセッサの内部メモリ）に登録することで、その決定された要求処理順序で、バッファ109に存在するI/O要求（或いはそれに含まれるコマンド又はデータ）の処理を行う。

**【0089】**

要求処理順序を決定する方法としては、様々な方法が考えられる。以下、図8以降を参照して、要求処理順序の決定方法の例について説明する。

**【0090】**

図8は、コマンド制御部102が、第1の要求処理順序決定方法でI/O要求の処理順序を決定するときの処理流れの概要を示す。

**【0091】**

コマンド制御部102は、I/O要求処理を開始した場合（ステップS81）、I/O要求バッファ109を参照し、そこにI/O要求が存在すれば（S82でY）、実行すべき（或いは処理可能な）I/O要求の数をチェックする（S83）。なお、「実行すべき（或いは処理可能な）I/O要求」とは、例えば、自分が有する全てのデータ又はコマンドがバッファ109に登録されており、且つ、それら全てのデータ又はコマンドのうち少なくとも1つのデータ又はコマンドが未処理の状態になったI/O要求である。

**【0092】**

さて、S83の結果、実行すべきものとして存在するI/O要求の個数がK個未満（ $K \geq 2$ ）である場合（S83でN）、コマンド制御部102は、バッファ109内のそのI/O要求を実行する（S88）。

**【0093】**

一方、S83の結果、実行すべきものとして存在するI/O要求の個数がK個以上Q個以下である場合（S83でY）、コマンド制御部102は、バッファ109内にある先着M個（ $K \leq M \leq Q$ ）のI/O要求の各々について、そのI/O要求に含まれるコマンド及びデータの属性と、性能情報テーブル1060（及び／又は統計情報テーブル1061）とに基づき、I/Oレスポンスタイムの期待値の合計を算出、換言すれば、各I/O要求のI/Oレスポンスタイムを予測する（S84）。そして、コマンド制御部102は、各I/O要求について算出された期待値の合計に基づいて、M個のI/O要求の平均レスポンスタイムが好適な平均レスポンスタイムになるものを識別し、識別された平均レスポンスタイムを得ることができるような、I/O要求の要求処理順序を決定する（S85）。コマンド制御部102は、決定した要求処理順序に基づいて、M個のI/O要求の各々の処理順位をI/O要求バッファ109に登録する。そして、コマンド制御部102は、M個のI/O要求を、その登録した処理順位で実行するための処理を開始し（S86）、M個のI/O要求について処理が終れば終了となる（S87）。

**【0094】**

なお、この図8に示した処理流れにおいて、例えば、上記実行すべきI/O要求の数がQ個に達した時点で、S81が開始されても良い。また、上記K及びQの値は、例えば、予め変更不能に或いは変更可能に決められておくパラメータである。K及びQの値がより大きくなると、処理手順のスケジューリング対象となるI/O要求の数が増えるので、M個のI/O要求の平均レスポンスタイムをより小さくすることが可能となるが、その反面、I/O実行順序を決定するためのオーバーヘッドが大きくなる。

**【0095】**

さて、図9は、図8の処理流れで示したS85における要求順序決定の一例を示す。

**【0096】**

コマンド制御部102は、M個のI/O要求について算出したM個の期待値合計のうち、最大のものと最小のものとを比較する（S851）。

**【0097】**

S851の比較の結果、上記最大と最小との差が所定の閾値S1より大きい場合（S851でY）、コマンド制御部102は、M個のI/O要求の要求処理順序を、I/Oレスポンスタイムの期待値合計が小さい順序に決定し、決定した順序をI/O要求バッファ109に書込む（S852）。

**【0098】**

一方、S851の比較の結果、上記最大と最小との差が所定の閾値S1以下の場合（S8

51でN)、コマンド制御部102は、M個のI/O要求の要求処理順序を、その受付順(例えば、Define Extentコマンドの受信順)と同じ順序に決定し、決定した順序をI/O要求バッファ109に書込む(S853)。

【0099】

S852又はS853の処理が終ると、S85の終了となる(S854)。

【0100】

なお、上述した説明において、所定の閾値S1は、例えば、予め変更可能に又は変更不可能に設定されるパラメータである。また、この閾値S1を設定せずに、上記M個の期待値合計のうち第1の期待値合計と第2の期待値合計と間の単なる大小関係だけで、要求処理順序をどのようにするかを制御するようにすれば、M個のI/O要求の要求処理順序についてきめ細かい制御が可能となるが、I/O要求の要求処理順序を、I/O要求の受付順序(例えば、Define

Extentコマンドの受信順序)とは違う順序に入れ替えることにより発生するオーバーヘッドを考慮に入れると、例えば、閾値S1を適切な値に設定し、I/Oレスポンスタイムの期待値合計がある範囲に収まる複数のI/O要求に関しては、I/O要求を受け付けた順序で実行することにより、前記のオーバーヘッドを削減することが可能になる。別の言い方をすれば、例えば、I/Oレスポンスタイムの期待値合計が所定の範囲内にある2以上のI/O要求については、それら2以上のI/O要求の受付順序と同じ順序で処理が行われるようにし、一方、I/Oレスポンスタイムの期待値合計が所定の範囲外にある別の2以上のI/O要求については、それら別の2以上のI/O要求の要求処理順序は、I/Oレスポンスタイムの期待値合計が小さい順序に決定されても良い。

【0101】

以上が、第1の要求処理順序決定方法に基づくコマンド制御部102の処理流れである。次に、図10以降を参照して、第2の要求処理順序決定方法に基づくコマンド制御部102の処理流れについて説明する。

【0102】

図10は、コマンド制御部102が、第2の要求処理順序決定方法でI/O要求の処理順序を決定するときの処理流れの概要を示す。

【0103】

コマンド制御部102は、或るI/O要求に対する処理を開始する(S101)。そして、コマンド制御部102は、当該I/O要求を開始した時刻をタイムスタンプとして記憶(例えば、バッファ109に記録)する。

【0104】

次に、コマンド制御部102は、当該I/O要求の処理を開始してから経過した時間をチェックする(S102)。このS102の処理は、所定のタイミング(例えば或る一定の周期)で行う。

【0105】

S102のチェックの結果、上記経過時間が予め定義されている経過時間閾値S2以下の場合は(S102でY)、コマンド制御部102は、当該I/O要求の処理を続行する(S106)。

【0106】

一方、S102のチェックの結果、上記経過時間が閾値S2を超えている場合(S103でN)、コマンド制御部102は、当該I/O処理を続行するか、或いは、他のI/O要求に処理を切り替えるかを判定する(S103)。別の言い方をすれば、コマンド制御部102は、現在処理中のI/O要求よりも直ちに処理することが望ましい別のI/O要求が存在するか否かをチェックする。

【0107】

S103において、処理を切り替えないと判定された場合は(S103でN)、コマンド制御部102は、現在実行中のI/O要求処理を続行する。

**【0108】**

一方、S103において、処理を切り替えると判定した場合（S103でY）、コマンド制御部102は、処理を切り替えて実行するI/O要求の要求処理順序を決定する（S104）。この要求処理順序は、例えば、図8及び図9を参照して説明した方法と同様に、実行すべきM個のI/O要求のレスポンスタイムの期待値合計を算出し、算出されたM個の期待値合計が小さい順に、上記M個のI/O要求の要求処理順序を決定することができる。

**【0109】**

その後、コマンド制御部104は、決定した要求処理順序に従って、実行すべき複数のI/O要求についての処理を開始する（S105）。

**【0110】**

以上が、第2の要求処理順序決定方法に基づくコマンド制御部102の処理流れの概要である。なお、この流れにおいて、閾値S2は、一つのI/O要求を続けて処理する時間長の閾値である。

**【0111】**

記憶制御装置10において、処理完了までに比較的長い時間を要する第1のI/O要求の処理中に、比較的短い時間で処理することが期待できる第2のI/O要求を受けた場合、第1のI/O要求の処理を長時間行くと、前述した3つ目の考察から分かるように、第1及び第2のI/O要求の平均レスポンスタイムが劣化してしまう。

**【0112】**

そこで、それを避けるため、第2の要求処理順序決定方法では、I/O要求の処理が開始されてからの経過時間の閾値S2を設け、その経過時間が閾値S2を越えた場合、コマンド制御部102によって、現在処理中のI/O要求よりも直ちに処理することが望ましい別のI/O要求が存在するか否かのチェックが行われる。これにより、長い処理時間を要する第1のI/O要求のために、処理時間が短くて済む第2のI/O要求のI/Oレスポンスタイムが劣化することを回避し、且つ、第1及び第2のI/O要求の平均レスポンスタイムが劣化することも回避する。

**【0113】**

図11は、図10に示した処理流れにおけるS103の具体的な処理流れを示す。

**【0114】**

コマンド制御部102は、I/O要求バッファ109内に、処理待ち状態のI/O要求があるか否かをチェックし、処理待ち状態のI/O要求が格納されていれば、性能情報テーブル1060（及び／又は統計情報テーブル1061）を用いて、当該1以上のI/O要求についてのI/Oレスポンスタイムの期待値合計を算出し、算出された期待値合計のうち、I/O要求バッファ109内に予め設定される閾値S3より小さい期待値合計を持つI/O要求があるか否かを判定する（S1031）。

**【0115】**

S1031において、閾値S3より小さな期待値合計となったI/O要求がある場合（S1031でY）、コマンド制御部102は、そのI/O要求を、要求処理順序を決定する際の要素となるI/O要求として選択し（S1032）、S103を終える（S1033）。

**【0116】**

一方、S1031において、I/O要求バッファ109内に処理待ち状態のI/O要求が無い場合、又は、上記閾値S3より小さな期待値合計となったI/O要求が無い場合、コマンド制御部102は、S103において、別のI/O要求の処理に切り替えはしないと決定して、S103を終える（S1033）。

**【0117】**

なお、この処理流れにおいて、閾値S3とは予め設定される閾値である。現在実行中のI/O処理が閾値S2を越える処理時間がかかっているとはいえ、やみくもに別のI/O要求に切り替えても、平均レスポンスタイムの短縮化は必ずしも図れない。現在実行中のI/O

O処理が閾値S2を超え、別のI/O要求の処理を切り替える場合、この切り替えたI/O要求の処理がある程度短い処理で終了する場合にのみ、複数のI/O要求の平均レスポンスタイムを向上させる目的にかなうと考えられる。そこで、図11を参照して説明した方法によって、現在実行中のI/O要求処理を別のI/O要求処理に変えるか否かが判断される。

#### 【0118】

以上、上述した実施形態によれば、受信された2以上のI/O要求に含まれる複数のコマンド要素（コマンド又はデータ）に基づいて、それら2以上のI/O要求の平均レスポンスタイムが、複数のコマンド要素又は2以上のI/O要求をその受信順序に従って処理するときの値以下になるような、受信順序とは異なる処理手順が決定され、それに基づいて、上記複数のコマンド要素の処理が開始される。これにより、2以上のI/O要求の平均レスポンスタイムを、単にそれらの受信順序に従って処理されるとき平均レスポンスタイムよりも短くすることができる。

#### 【0119】

なお、上述した説明では、上位装置20から発行されたI/O要求を受信するチャンネル制御部101、I/O要求バッファ109及びコマンド制御部102について詳細に説明したが、他のチャンネル制御部104、I/O要求バッファ110及びコマンド制御部105も実質的に同様に機能する。

#### 【0120】

また、上述した実施形態において、例えば、コマンド制御部102は、様々なタイミングで（例えば、キャッシュヒットしたとき及びキャッシュミスしたときの少なくとも一方のときに）、要求処理順序を決定（或いは変更）する処理を開始しても良い。

#### 【0121】

また、上述した実施形態において、所定の種々の値（例えば、K、Q、及び閾値S1～S3の少なくとも一つ）は、処理対象のI/O要求の切り替えの際のタイムロス等の種々の情報に基づいて適切な値に設定される。また、その値は、例えば、一旦設定されて記憶制御装置10が起動した後は、変更が不可能であっても良いし、或いは、記憶制御装置10内のチューニング等によって変更が可能であっても良い。

#### 【0122】

また、上述した実施形態において、例えば、要求処理順序を決定する、記憶制御装置10における構成要素は、コマンド制御部102、105に限らず他の構成要素が行っても良い。

#### 【0123】

また、上述した実施形態において、例えば、コマンド制御部102は、或るI/O要求の処理中に、他のI/O要求の処理に切り替えるか否かの判断を行うときは、その或るI/O要求については、その処理完了までにかかると推測される残りの時間長に基づいて、上記判断が行われても良い。また、例えば、要求処理順序の決定が行われるときには、その残りの時間長に基づいて、当該決定が行われても良い。なお、その残り時間長は、例えば、コマンド制御部102が、処理中のI/O要求において、未処理の1以上のコマンド要素の属性に基づいて1以上のコマンド要素の期待値合計を算出することで求めても良い。

#### 【0124】

また、上述した実施形態では、I/O要求単位で処理手順が決定され、I/O要求単位で一つの処理が行われるが、例えば、それに代えて、I/O要求のコマンド要素単位で処理順序が決定されてコマンド要素の処理が行われても良い。具体的には、例えば、第1のI/O要求に、第1及び第2のコマンド要素が含まれていて、第2のI/O要求に、第3及び第4のコマンド要素が含まれている場合、上記実施形態では、第1及び第2のコマンド要素の処理が行われてから、第3及び第4のコマンド要素が行われるが、この変形例では、第1、第3、第2、第4の順でコマンド要素が処理されるような要求処理順序が決定されても良い。この方式は、2つ目の考察からすれば、平均レスポンスタイムの向上につ

なかりにくいこともあるが、例えば、或る I/O 要求について処理時間長が膨大なコマンド要素の処理中に、他の I/O 要求について処理時間が極小のコマンド要素を受信した場合、その極小のコマンド要素を先に処理してしまった方が、平均レスポンスタイムを短縮化することができる。

#### 【0125】

なお、この変形例における更なる変形例として、情報セット単位（例えば I/O 要求単位）で要求処理順序を決定する第 1 の情報処理モードと、情報要素単位（例えばコマンド要素単位）で要求処理順序を決定する第 2 の情報処理モードとがあり、これら 2 つの情報処理モードが、所定のモード選択方法に基づいて選択的に使用されても良い。具体的には、例えば、情報セットを発行する上位装置 20 が記憶制御装置 10 に所望の情報処理モードを通知することで、記憶制御装置 10 が、その上位装置に対してはその通知された情報処理モードに基づいて要求処理順序を決定するようにしても良い。また、例えば、コマンド制御部 102 は、バッファ 109 に、全ての情報要素（コマンド要素）が記録された情報セット（I/O 要求）が複数個ある場合には、その複数の情報セットについては情報セット単位で処理順序を決定し、一部の情報要素しか記録されていない別の情報セットが複数個ある場合には、その複数の別の情報セットについては情報要素単位で処理順序が決定されても良い。また、例えば、記憶制御装置 10 は、処理対象の I/O 要求の切り替えの際にロスする時間長情報等、平均レスポンスタイムの向上に関わる有用な種々を情報を記憶しておき、その情報に基づいて、任意タイミングで（例えば、或る情報セット又は或る情報要素の処理を開始してから一定時間を経過したとき）、且つ、任意の要求処理手順決定方法で、2 以上の情報セット又はそれらに含まれる複数の情報要素の要求処理手順を決定しても良い。なお、情報要素単位で要求処理順序が決定された場合には、各情報要素毎に、順位がバッファ 109 に記録されても良い。また、決定される要求処理手順は、例えば、情報セット又は情報要素の受信順序とは異なる順序であって、その受信順序のときの平均レスポンスタイムよりも短い平均レスポンスタイムになる順序である。

#### 【0126】

以上、本発明の好適な実施形態及び変形例を説明したが、これらは本発明の説明のための例示であって、本発明の範囲をこの実施形態及び変形例にのみ限定する趣旨ではない。本発明は、他の種々の形態でも実施することが可能である。

#### 【0127】

例えば、上記期待値は、レスポンスタイムの期待値ではなく、記憶制御装置 10 において行われる処理に要する時間長の期待値であっても良い。また、インターロックを取らずに上位装置がパイプライン的にデータ又はコマンドを送信することができる通信プロトコル（一例として FC-SB2 プロトコル）は、別の言い方をすれば、例えば、上位装置と記憶サブシステム（記憶装置 60 及び記憶制御装置 10 を含むシステム）とが、記憶サブシステム側からの応答とは非同期に、上位装置が複数のコマンド及びデータのチェーンで構成される I/O 要求を記憶システムに対して発行することができるプロトコルである。

#### 【0128】

また、例えば、本発明の第 1 の好適な実施形態では、前記決定手段は、前記複数の情報要素に基づいて、前記 2 以上の情報セットの各々について、処理に要すると期待される時間長に関する期待値を取得し（例えば算出し）、前記取得された期待値に基づいて、前記処理手順を決定する。

#### 【0129】

また、例えば、本発明の第 2 の好適な実施形態では、上記第 1 の好適な実施形態において、情報処理装置は、情報要素に関する複数の情報要素属性の各々について、その情報要素属性を持つ情報要素の処理に要すると期待される時間長に関するサブ期待値を表すサブ期待値情報を記憶するサブ期待値記憶手段を更に備える。この場合、前記決定手段が、前記複数の情報要素の各々の情報要素属性と、前記記憶されているサブ期待値情報とに基づいて、前記複数の情報要素の各々についての前記サブ期待値を取得し、前記取得されたサブ期待値を用いて、前記 2 以上の情報セットの各々についての前記期待値を取得する。な

お、「情報要素属性」には、例えば、情報要素の種類、データサイズ及び発行元のうちの少なくとも1つが含まれる。

【0130】

また、例えば、本発明の第3の好適な実施形態では、上記第2の好適な実施形態において、前記決定手段は、前記取得されたサブ期待値が小さい情報要素又は情報セットほど早い順番で処理されるような前記処理手順を決定する。

【0131】

また、例えば、本発明の第4の好適な実施形態では、上記第1の好適な実施形態において、前記決定手段は、前記情報処理手段によって過去に処理された情報要素の属性とその情報要素の処理に要した時間長との履歴に関する情報に基づいて、前記2以上の情報セットの各々についての前記期待値を取得する。

【0132】

また、例えば、本発明の第5の好適な実施形態では、前記決定手段は、前記複数の情報要素についての複数通りの処理手順にそれぞれ対応した複数通りの前記平均に関する値のうち最小の値となるような処理手順を決定する。

【0133】

また、例えば、本発明の第6の好適な実施形態では、前記決定手段は、前記2以上の情報セットについて、前記処理時間長の最小と最大との差分を取得し、前記差分が所定値以上である場合に、前記処理手順を決定する処理を実行する。

【0134】

また、例えば、本発明の第7の好適な実施形態では、前記複数の情報要素に、前記情報処理手段によって処理中の情報要素が含まれているときは、前記決定手段は、その情報要素の処理に要する処理時間長のうちの残りの時間長に基づいて、前記処理手順を決定する。

【0135】

また、例えば、本発明の第8の好適な実施形態では、情報処理装置は、前記情報セットについての処理が完了する場合に、その処理が終了したことを意味する終了信号をその情報セットの発行元の情報セットソースに送信する終了信号送信手段を更に備える。この場合、前記情報セットについての前記処理時間長は、前記情報セットソースが前記情報セットを発行してから前記終了信号を受信するまでのレスポンスタイムに関する時間長である。

【0136】

また、例えば、本発明の第9の好適な実施形態では、前記決定手段は、前記受信された2以上の情報セットについての受信状況が所定の状況である場合に（例えば、受信した情報セットの数が所定の数以上である場合に）、前記処理手順を決定する処理を実行する。なお、その場合でなければ、例えば、前記決定手段は、前記情報処理手段が、現在処理中の情報セット又は情報要素の処理を継続して実行する、或いは、受信した順序で情報セット又は情報要素の処理を行うようにする。

【0137】

また、例えば、本発明の第10の好適な実施形態では、前記決定手段は、前記情報セット又は前記情報要素についての処理が開始されてから所定時点までの経過時間長を把握し、前記経過時間長が一定の時間長を超えたときに、前記処理手順を決定する処理を実行する。具体的には、例えば、前記決定手段は、上記経過時間長が前期一定の時間長を超えた1又は複数の情報セット又は2以上の情報要素について上記期待値を取得し、取得した期待値に、所定の閾値より小さい期待値がある場合には、前記処理手順を決定する処理を実行する。なお、それらの条件を満たさない場合、例えば、前記決定手段は、前記情報処理手段が、現在処理中の情報セット又は情報要素の処理を継続して実行する、或いは、受信した順序で情報セット又は情報要素の処理を行うようにする。

【0138】

また、例えば、本発明の第11の好適な実施形態では、情報処理装置は、物理的又は論

理的な 1 又は複数の記憶デバイスと通信可能に接続することができる。受信手段は、1 以上の情報要素を有する I/O 要求を発行する 1 又は複数の I/O 要求ソースから前記 I/O 要求の各々に含まれる情報要素を受信する。情報処理手段は、前記受信された I/O 要求に含まれている情報要素に基づいて、前記記憶デバイスからデータをリードし I/O 要求ソースに送信する、又は、前記 I/O 要求ソースからのデータを前記記憶デバイスにライトする処理を行う。決定手段は、受信された 2 以上の I/O 要求に含まれる未処理又は処理中の複数の情報要素に基づいて、前記 2 以上の I/O 要求又は前記複数の情報要素の処理手順を決定するものであって、前記 2 以上の I/O 要求のレスポンスタイムの平均に関する値が、前記複数の情報要素又は前記 2 以上の I/O 要求をその受信順序に従って処理するときの値以下になるような、前記受信順序とは異なる前記処理手順を決定する。

【図面の簡単な説明】

【0139】

【図 1】本発明の一実施形態に係る情報処理装置が適用された記憶システムの全体構成を示す。

【図 2】ライト I/O 要求に対する処理シーケンスの一例。

【図 3】複数 I/O 要求の処理におけるレスポンスタイムの第 1 の考察。

【図 4】複数 I/O 要求の処理におけるレスポンスタイムの第 2 の考察（平均レスポンスタイムの考察）。

【図 5】複数 I/O 要求の処理におけるレスポンスタイムの第 3 の考察（平均レスポンスタイム）。

【図 6】共有メモリ 106 における情報性能テーブル 1060 の一例。

【図 7】I/O 要求バッファ 109 を示す図。

【図 8】コマンド制御部 102 が、第 1 の要求処理順序決定方法で I/O 要求の処理順序を決定するときの処理流れの概要を示す。

【図 9】図 8 の処理流れで示した S85 における要求順序決定の一例を示す。

【図 10】コマンド制御部 102 が、第 2 の要求処理順序決定方法で I/O 要求の処理順序を決定するときの処理流れの概要を示す。

【図 11】図 10 に示した処理流れにおける S103 の具体的な処理流れを示す。

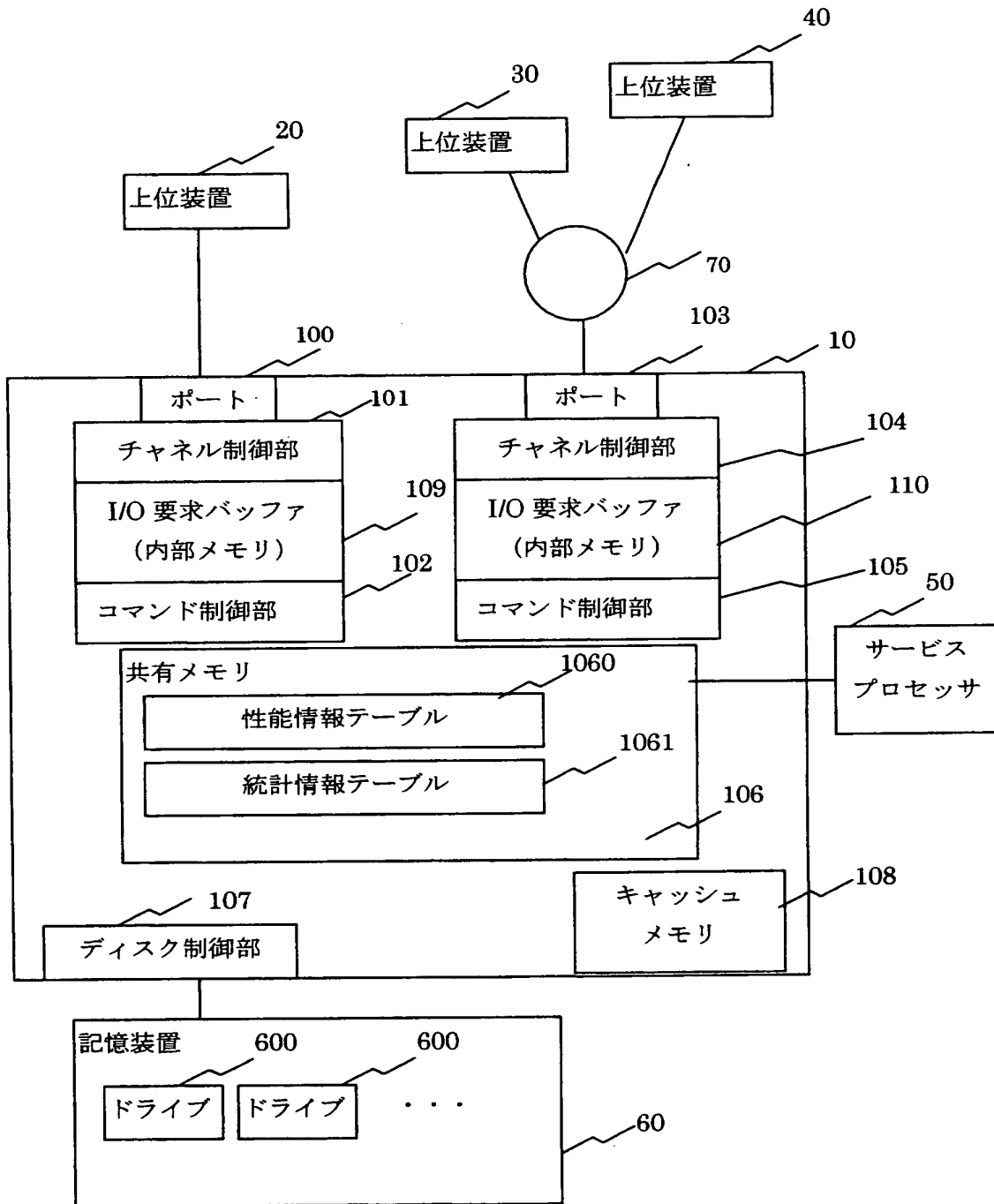
【符号の説明】

【0140】

- 10…記憶制御装置
- 20…上位装置
- 30…上位装置
- 40…上位装置
- 50…サービスプロセッサ
- 60…記憶装置
- 70…ダイレクタ
- 100、103…ポート
- 101、104…チャネル制御部
- 102、105…コマンド制御部
- 109、110…I/O 要求バッファ
- 106…共有メモリ
- 107…ディスク制御部
- 108…キャッシュ制御部
- 600、601…ドライブ
- 1060…性能情報テーブル
- 1061…統計情報テーブル

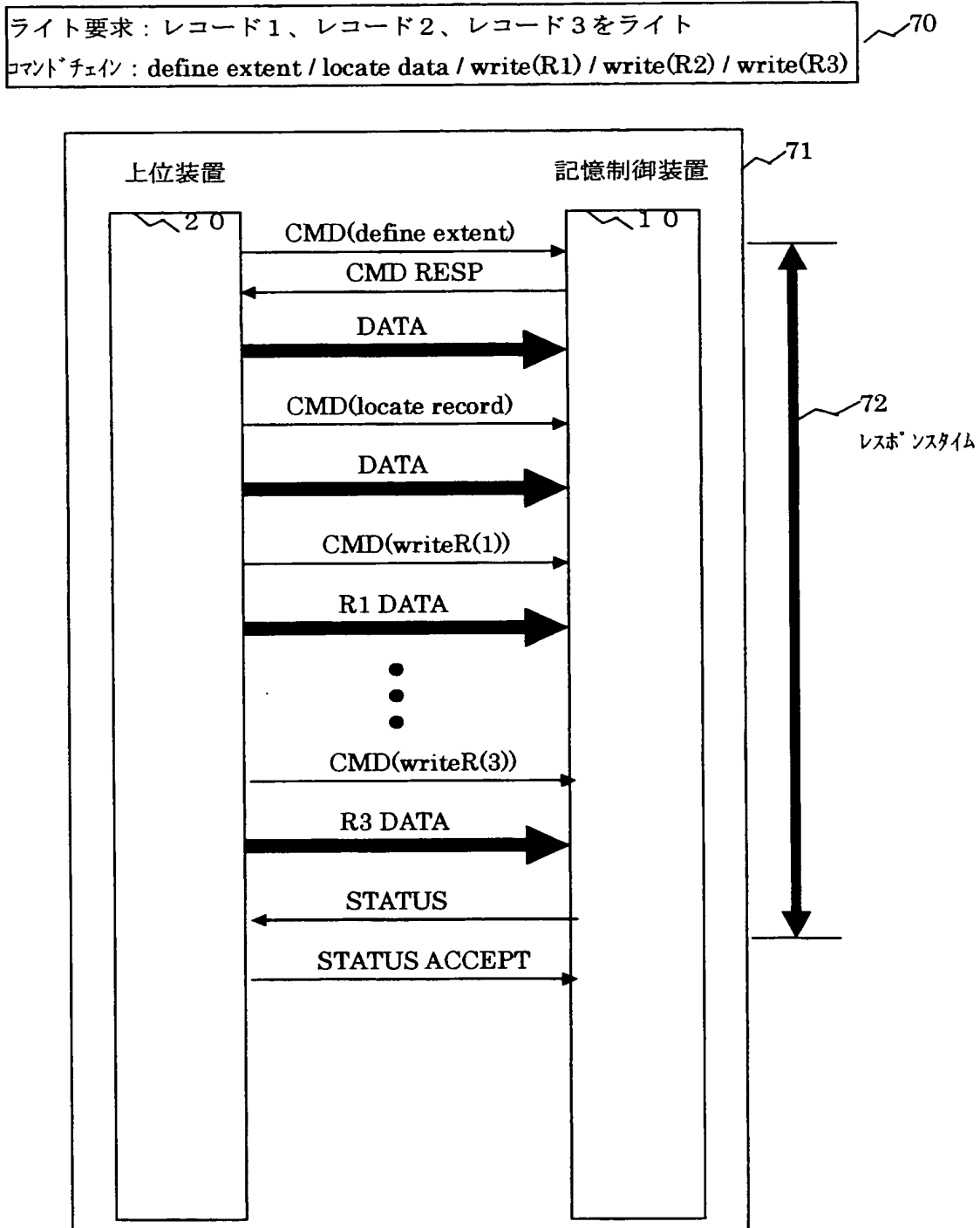
【書類名】 図面  
【図 1】

本発明の一実施形態に係る情報処理装置が適用されたシステム



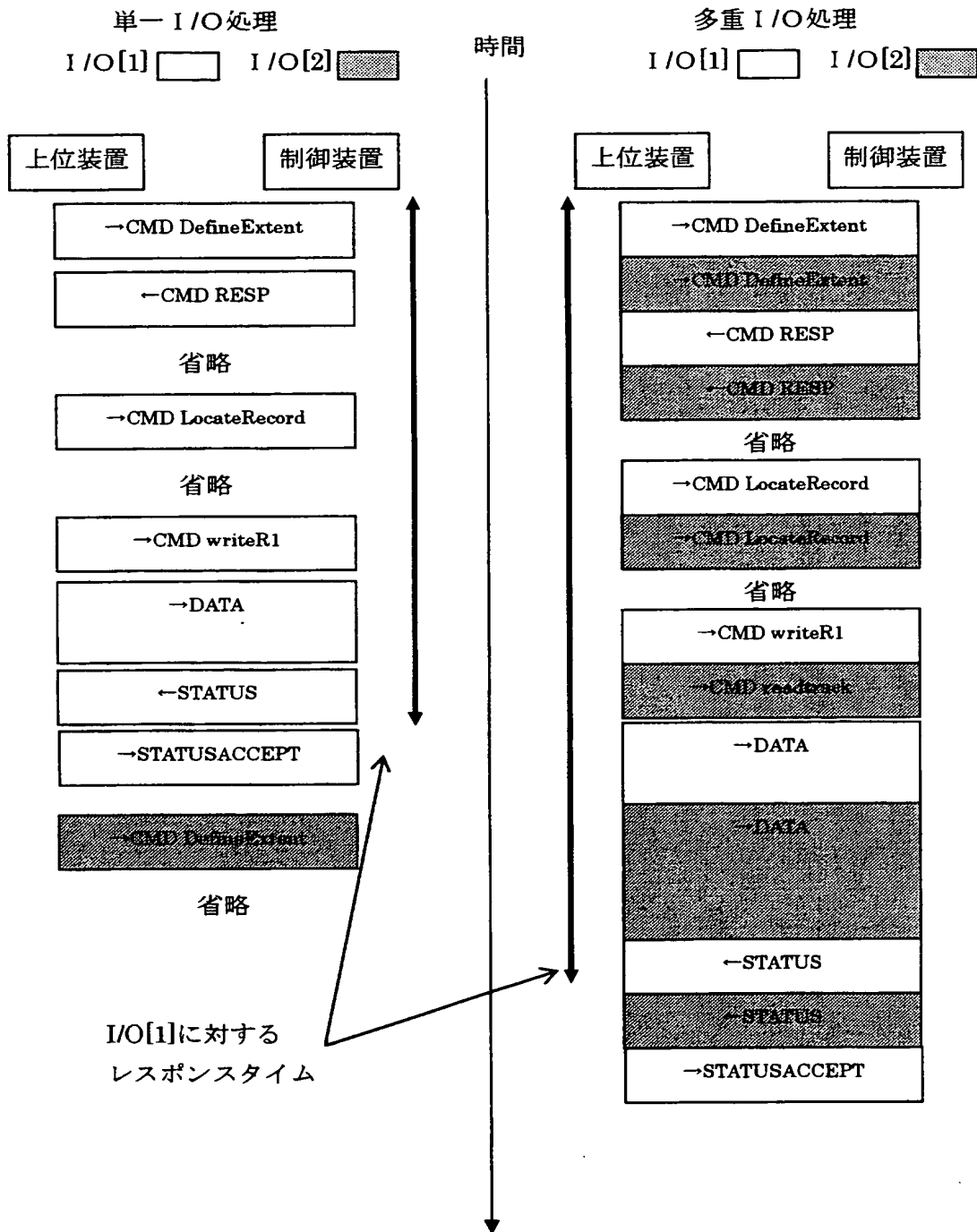
【図 2】

ライト I/O 要求に対する処理シーケンス



【図 3】

複数の I/O 要求の処理におけるレスポンスタイムの考察-1(単一 I/O 処理と多重 I/O 処理)

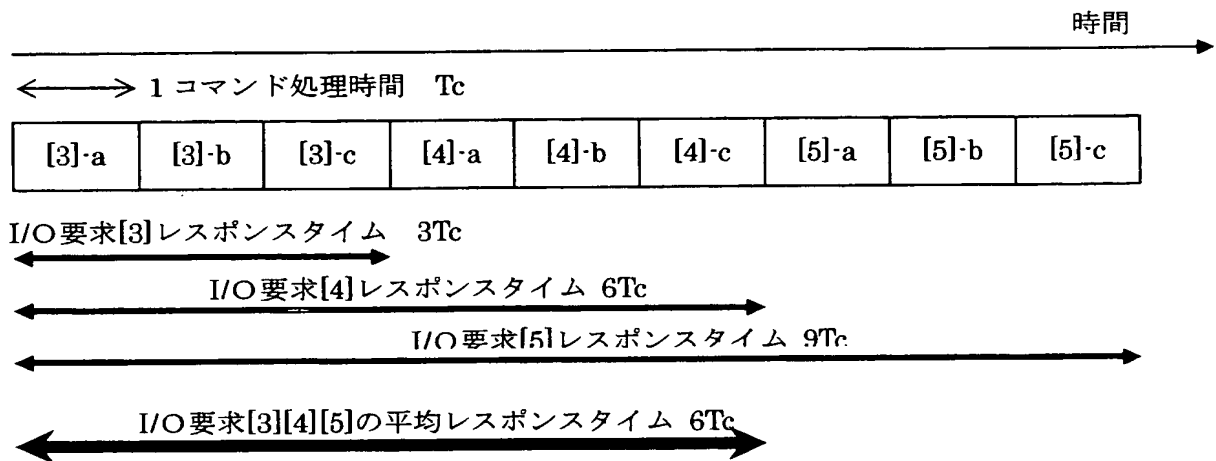


【図 4】

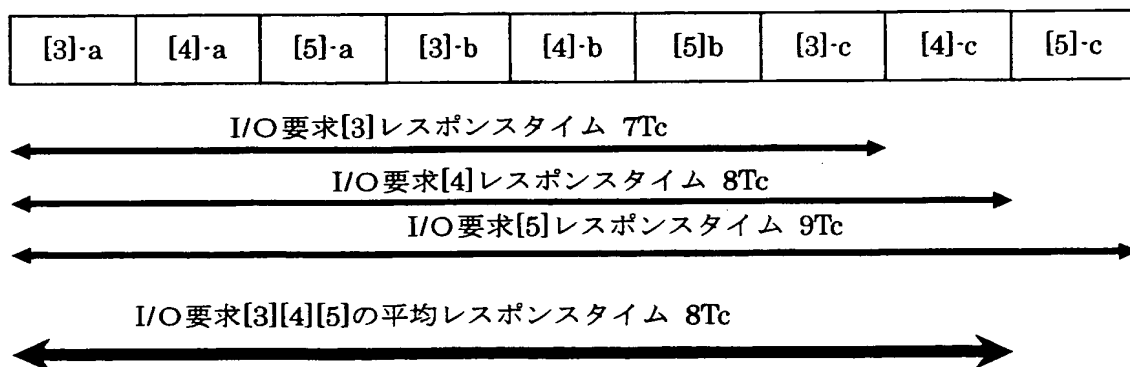
複数 I/O 要求の処理におけるレスポンスタイムの考察-2  
(シーケンシャルに処理/多重で処理)

I/O 要求[3]	コマンド[3]・a	コマンド[3]・b	コマンド[3]・c
I/O 要求[4]	コマンド[4]・a	コマンド[4]・b	コマンド[4]・c
I/O 要求[5]	コマンド[5]・a	コマンド[5]・b	コマンド[5]・c

ケース 3-A: I/O 要求毎にシーケンシャルに処理する



ケース 3-B: 複数 I/O のコマンドを混在させて処理する



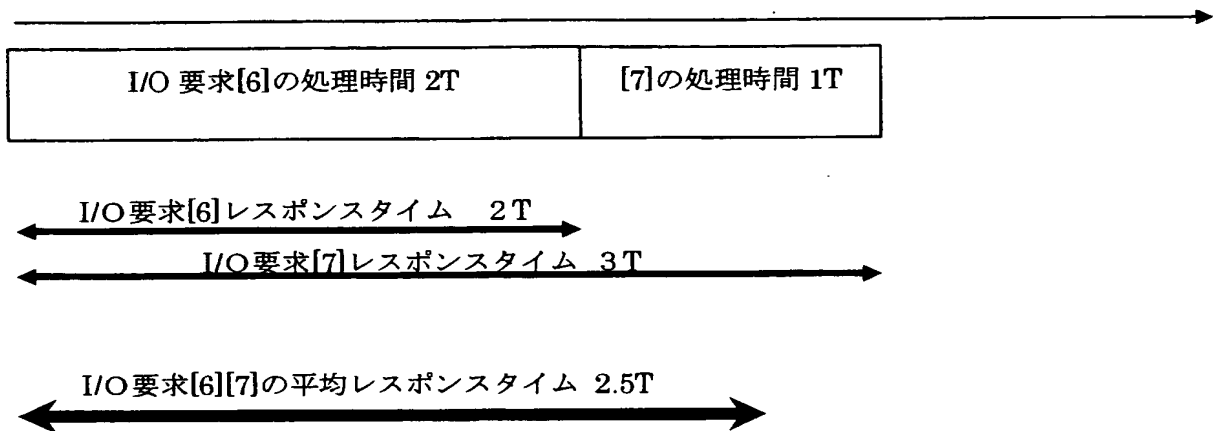
【図 5】

複数 I/O 要求の処理におけるレスポンスタイムの考察-2  
(処理順序)

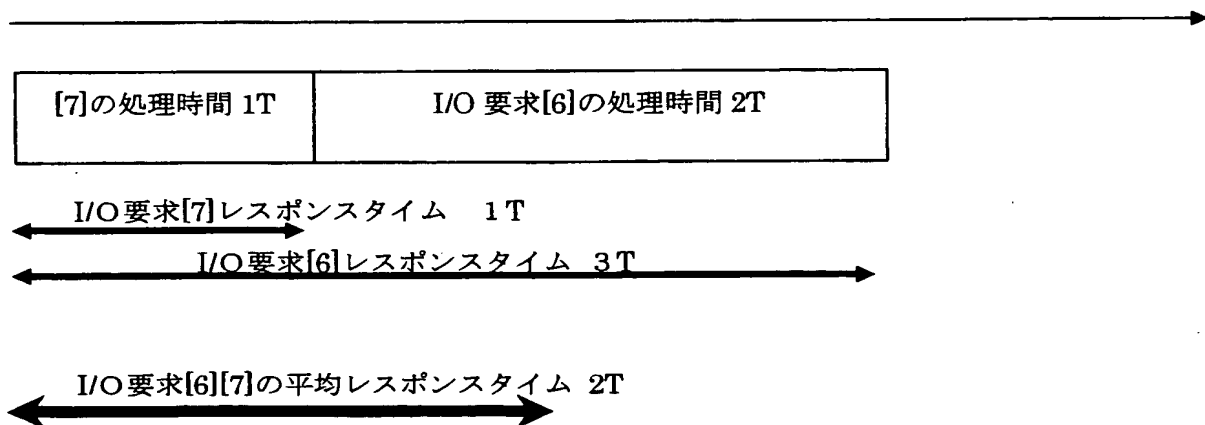
I/O 要求[6] 処理にかかる時間  $2T$

I/O 要求[7] 処理にかかる時間  $1T$

ケース 4-A: I/O 要求[6][7]の順に処理する



ケース 4-B: I/O 要求[6][7]の順に処理する



【図 6】

共有メモリ 106 内のテーブル仕様

1060

コマンド処理時間	
コマンド種別	処理時間の期待値( $\mu$ sec)
コマンド[1]	T1
コマンド[2]	T2
コマンド[3]	T3
⋮	⋮
⋮	⋮
⋮	⋮
コマンド n	Tn

1060A

データ転送処理時間		
転送種別/転送長(byte)		処理時間の期待値( $\mu$ sec)
リード 処理	データ長 a	Ta
	データ長 b	Tb
	⋮	⋮
	⋮	⋮
	⋮	⋮
ライト 処理		Tx
	データ長 a	Ta
	データ長 b	Tb
	⋮	⋮
	⋮	⋮
	データ長 x	Tx

1060B

【図 7】

I/O 要求バッファ

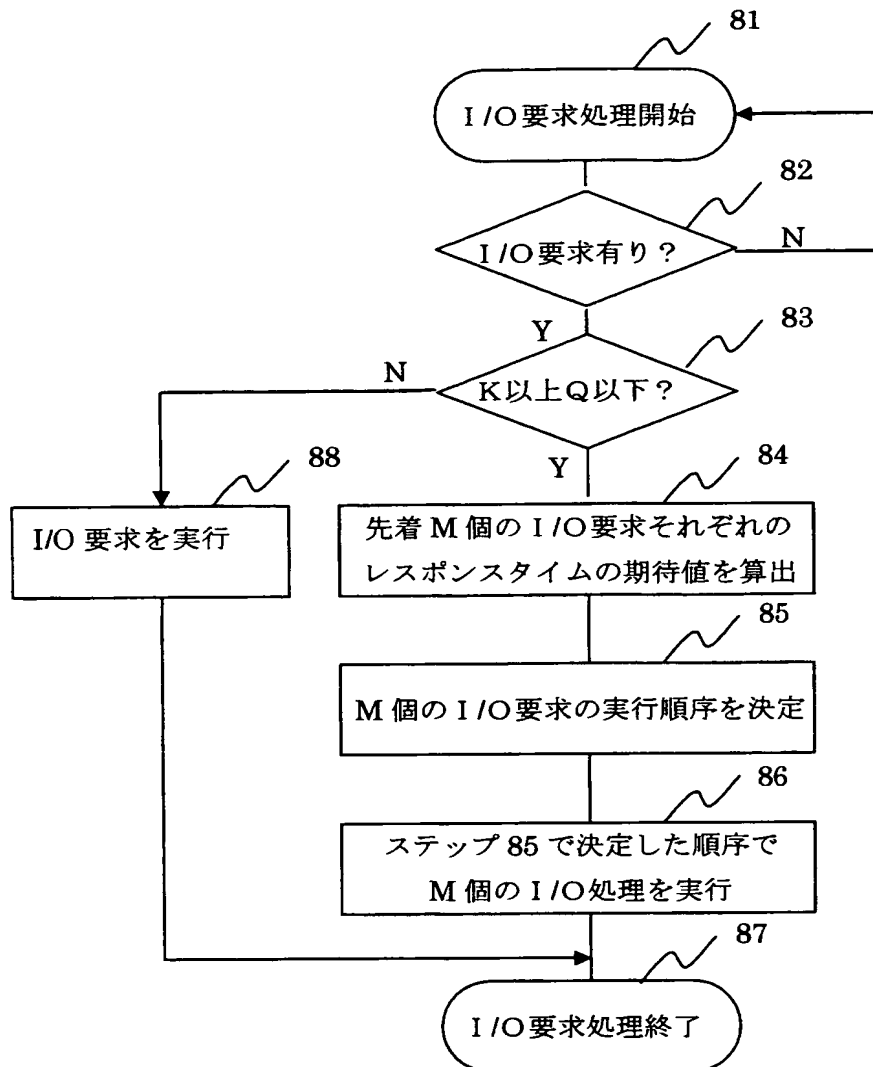
109,110

I/O 要求 1	I/O 要求 2	I/O 要求 3	I/O 要求 4		I/O 要求 N
実行順序情報	実行順序情報	実行順序情報	実行順序情報		実行順序情報
コマンド a	コマンド a	コマンド a	-		-
データ a	データ a	データ a	-		-
コマンド b	コマンド b	コマンド b	-		-
データ b	データ b	データ b	-		-
コマンド c	-	コマンド c	-		-
データ c-1	-	データ c-1	-		-
データ c-2	-	データ c-2	-		-
データ c-3	-	-	-		-
-	-	-	-		-
-	-	-	-		-
-	-	-	-		-

登録 I/O 数

【図 8】

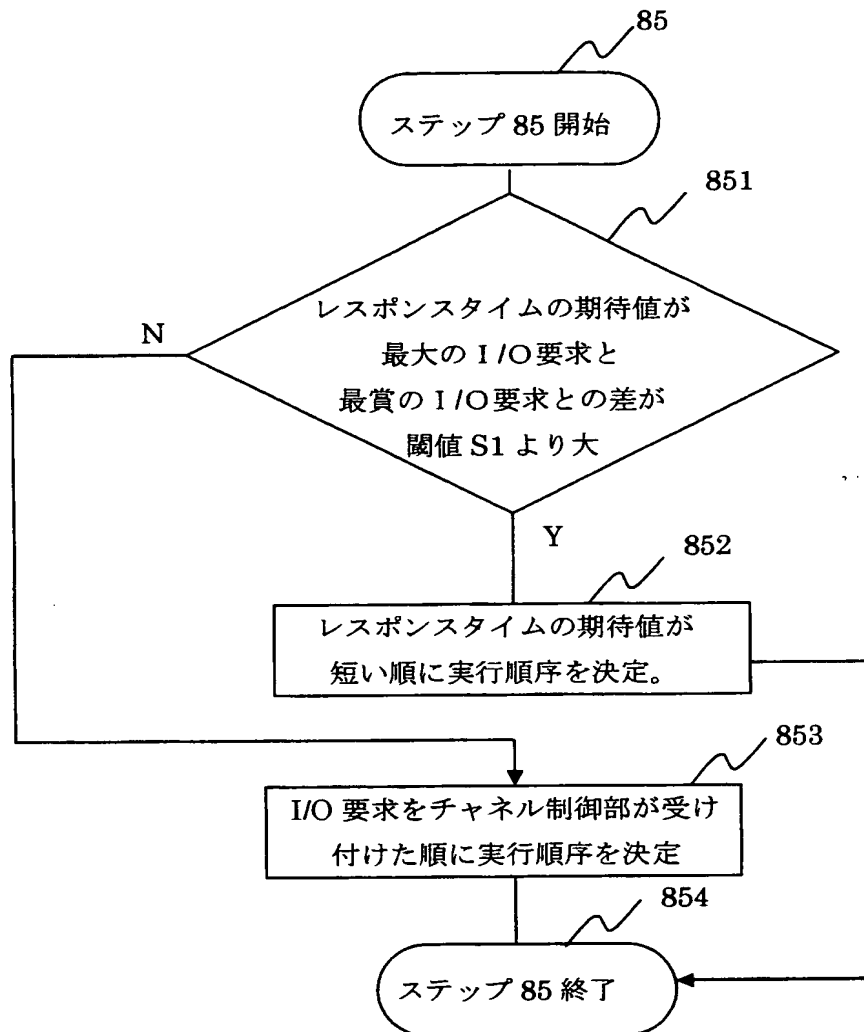
コマンド制御部における I/O 要求実行順序決定処理フロー(1)



【図9】

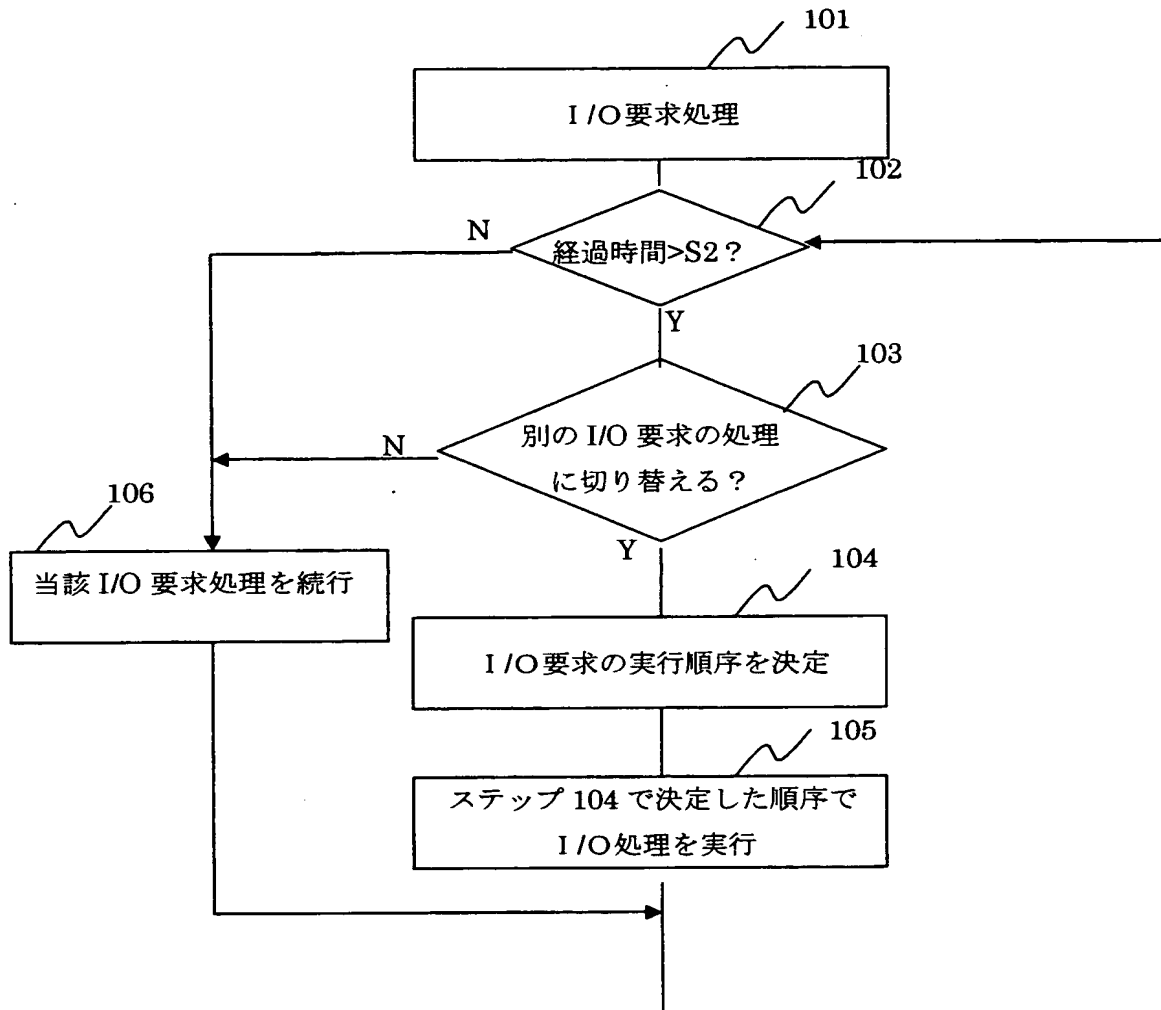
コマンド制御部における I/O 要求実行順序決定処理フロー(1)

ステップ 85 の詳細



【図 10】

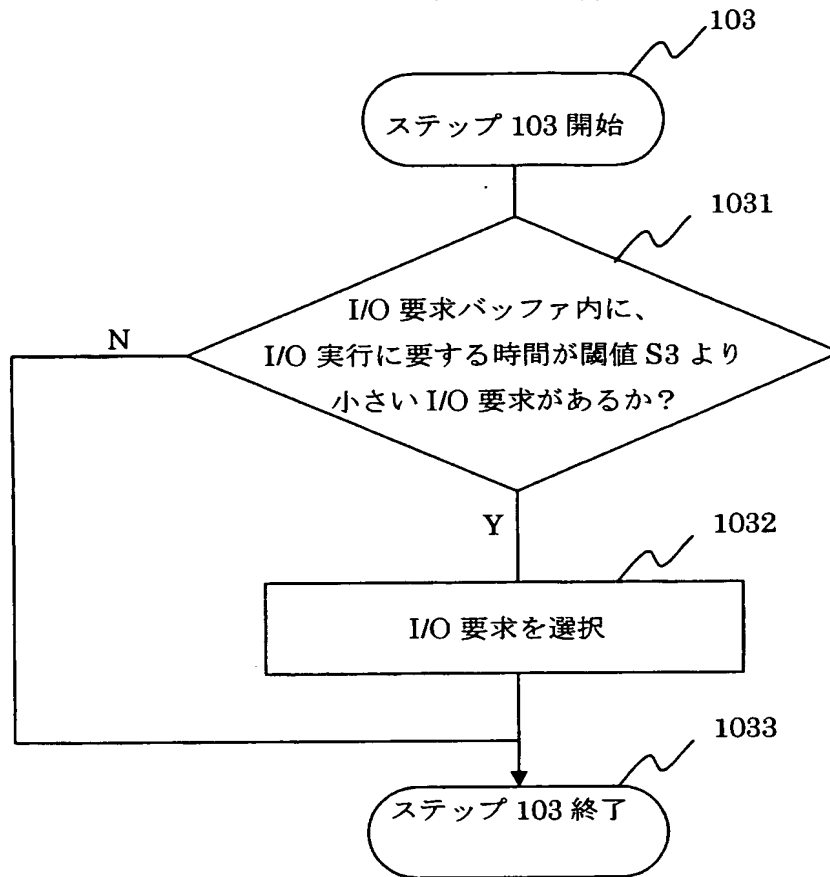
コマンド制御部における I/O 要求実行順序決定処理フロー(2)



【図 11】

コマンド制御部における I/O 要求実行順序決定処理フロー(2)

ステップ 103 の詳細



【書類名】 要約書

【要約】

【課題】 複数の情報セットの処理に要する処理時間長の平均に関する値を小さくする。

【解決手段】

情報処理装置（10）は、1以上の情報要素を有する情報セットを発行する1又は複数の情報セットソース（20）から情報セットの各々に含まれる情報要素を受信する受信手段（101）と、受信された情報要素の処理を行う情報処理手段（102）と、受信された2以上の情報セットに含まれる未処理又は処理中の複数の情報要素に基づいて、2以上の情報セット又は複数の情報要素の処理手順を決定するものであって、2以上の情報セットの処理時間長の平均に関する値が、複数の情報要素又は2以上の情報セットをその受信順序に従って処理するときの値以下になるような、受信順序とは異なる処理手順を決定する決定手段（102）とを備え、その決定された処理手順に基づいて複数の情報要素の処理を開始する。

【選択図】 図1

認定・付加情報

特許出願の番号	特願 2 0 0 3 - 3 4 7 9 4 2
受付番号	5 0 3 0 1 6 6 7 0 3 2
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 5 年 1 0 月 8 日

< 認定情報・付加情報 >

【提出日】 平成 15 年 10 月 7 日

特願 2 0 0 3 - 3 4 7 9 4 2

出 願 人 履 歴 情 報

識別番号

[ 0 0 0 0 0 5 1 0 8 ]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☒ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**